1986-09

A multiple linear regression model for predicting zone A retention by military occupational specialty.

Higham, Ronald P.

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A MULTIPLE LINEAR REGRESSION MODEL
FOR PREDICTING ZONE A RETENTION
BY MILITARY OCCUPATIONAL SPECIALTY

by

Ronald P. Higham

September 1986

Thesis Advisor: Jack B. Gafford
Co-advisor: Donald R. Barr

ECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
1 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (*If applicable*) Code 55	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (*City, State, and ZIP Code*) Monterey, California 93943-5000	7b. ADDRESS (*City, State, and ZIP Code*) Monterey, California 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (*City, State, and ZIP Code*)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (*Include Security Classification*)
A MULTIPLE LINEAR REGRESSION MODEL FOR PREDICTING ZONE A RETENTION BY MILITARY OCCUPATIONAL SPECIALTY

12 PERSONAL AUTHOR(S)
Higham, Ronald P.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (*Year, Month, Day*)	15 PAGE COUNT
Master's Thesis	FROM _____ TO _____	1986 September	75

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	First Term Reenlistment, First Term Retention Selective Reenlistment Bonus

19 ABSTRACT (*Continue on reverse if necessary and identify by block number*)

The Selective Reenlistment Bonus (SRB) program is designed to offer an attractive reenlistment incentive to improve manning in critical skills. To efficiently manage the SRB program, a requirement exists to maintain MOS level estimating factors for use in projecting retention rate improvement as a function of SRB award level. This thesis formulates and solves a mathematical model which explains the variation in zone A retention rates as a function of SRB award level and other factors believed significant in the reenlistment decision.

To allow for comparison of the estimating factors associated with the SRB variable across MOS, an overall projection model was developed. Stepwise multiple linear regression analysis techniques were used on a subset of the enlisted MOS inventory in the model development phase of this analysis. The proposed overall model was then fitted to a second subset of MOS to validate the assumptions and effectiveness of the proposed linear model.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL Jack B. Gafford	22b TELEPHONE (*Include Area Code*) (408) 646-3452	22c OFFICE SYMBOL Code 55Gf

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Approved for public release; distribution is unlimited.

A Multiple Linear Regression Model for Predicting Zone A Retention by Military Occupational Specialty

by

Ronald. P. Higham
Captain, United States Army
B.S., United States Military Academy, 1978

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1986

ABSTRACT

The Selective Reenlistment Bonus (SRB) program is designed to offer an attractive reenlistment incentive to improve manning in critical skills. To efficiently manage the SRB program, a requirement exists to maintain MOS level estimating factors for use in projecting retention rate improvement as a function of SRB award level. This thesis formulates and solves a mathematical model which explains the variation in zone A retention rates as a function of SRB award level and other factors believed significant in the reenlistment decision.

To allow for comparison of the estimating factors associated with the SRB variable across MOS, an overall projection model was developed. Stepwise multiple linear regression analysis techniques were used on a subset of the enlisted MOS inventory in the model development phase of this analysis. The proposed overall model was then fitted to a second subset of MOS to validate the assumptions and effectiveness of the proposed linear model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The Commander, United States Army Military Personnel Center (MILPERCEN), is responsible for developing and issuing policies, standards and procedures in the administration of the Selective Reenlistment Bonus (SRB) program. The SRB program is designed to offer an attractive reenlistment incentive to improve manning in the most critical skills. A primary consideration in the management of the SRB program is the historic effectiveness of an SRB in improving retention in a particular skill. In this study, the problem of measuring the historic effectiveness of the SRB program is modelled and solved using stepwise and ordinary least squares multiple linear regression analysis.

A. PROBLEM STATEMENT

The Commander, MILPERCEN must recommend to the Deputy Chief of Staff for Personnel (DCSPER) those Military Occupational Specialties (MOS) which should be included in the SRB program. The criteria used to determine which MOS should be included in the SRB program are outlined in the form of several guidelines (specifically, Title 37 United States Code, section 308, Department of Defense (DOD) Directive 1304.21 and DOD Directive 1304.22). Some criteria, such as replacement training costs, are easily quantified. Other criteria, such as the relative *unattractiveness* of each MOS compared to other military and civilian skills, are much more subjective.

One criterion upon which the decision to include a particular MOS in the SRB program is based is the projected improvement in retention in response to the bonus awarded. There must be a reasonable prospect of enough improvement in retention to justify the projected cost of the bonus. Therefore, a requirement exists to maintain estimating factors for use in projecting retention rate improvement as a function of SRB award level. DOD directs that these factors be developed from actual experience under the SRB program.

The improvement factors currently available are outdated and were developed without consideration to certain variables believed critical to an accurate projection of retention at the MOS level.

B. BACKGROUND

In September 1981, the DCSPER requested that the Commander, United States Army Concepts Analysis Agency (CAA) establish a study group to develop an improved methodology for allocation of SRB funds. An intermediate goal of the study group was to quantify the effect of SRB on retention; that is, develop a set of historically based improvement factors. These factors were to replace similar improvement factors published by the Rand Corporation in September 1977 [Ref. 1]. The DCSPER suggested that the Rand factors were no longer valid, in light of more recent trends in retention, pay and civilian perception of military service.

In August 1982, the study was completed by CAA. Included in their final report [Ref. 2] were a set of MOS and reenlistment zone specific SRB effectiveness factors. These factors were said to represent the net change in retention rate for a given MOS brought on by a change in the SRB authorized that MOS. The factors were actually the estimated regression coefficients of the carrier variable SRB in the multiple linear regression model used to explain retention rate behavior for all MOS during the previous five years. The specific model follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_3^2 \qquad (1.1)$$

$$+ \beta_5 X_3^3 + \alpha_1 Z_1 + \alpha_2 Z_2 + \varepsilon$$

where:

Y = retention rate
X_1 = SRB multiplier
X_2 = year
X_3 = calender quarter
Z_1 = unemployment rate
Z_2 = Consumer Price Index
ε = error component with assumed distribution $N(0, \sigma^2)$.

While the study group cautioned against using the retention improvement factors (estimated regression coefficient b_1) for longer than two years, no provisions were made for the periodic re-estimation of those coefficients. Hence, the current set of coefficients are a function of data which are at least five years old. Additionally, while diagnostics from the CAA model support a reasonably good fit to the data available,

no attempt was made by the CAA analysts to account for the effects of factors such as population demographics and promotion opportunity.

The Deputy Chief of Staff for Plans (DCSPLANS), MILPERCEN submitted this problem, with the below stated objectives, to the Naval Postgraduate School, pursuant to a special thesis study / management program. Under this program, a participating Army student works with MILPERCEN to resolve a current problem and receives a follow-on assignment to the Personnel Center upon graduation. All research costs and other costs associated with thesis preparation are borne by MILPERCEN.

C. STUDY OBJECTIVE

The objective of this study is to formulate a mathematical model which explains the variation in zone A enlisted retention rates over time at the MOS level of detail. Variables representing promotion opportunity to grades E5 and E6 and a variable representing SRB award level are to be considered as candidate explanatory variables.

D. MODEL AND SOLUTION APPROACH

The mathematical formulation proposed in this study is an ordinary least squares multiple linear regression model with higher order terms. It is our intention to carefully select our dependent and independent variables so that the model can be used in a predictive manner: given a set of outcomes on the explanatory variables, we wish to predict an outcome on our selected response variable with a measureable degree of precision.

Our objective is to build a model which can predict zone A retention at the MOS level. It is likely therefore, that if each MOS subpopulation were studied independently, the carrier variables included in the final model (selected by some system of rules) would not be identical for each MOS. This situation, for our purposes, is not acceptable.

The intentions of our user dictate that we select a best model and apply it for all MOS. As has already been mentioned, the SRB managers have used the estimated coefficient of the carrier variable SRB (we refer to this estimate henceforth simply as b_1) to compare the effects on retention of varying the SRB level across several, or even all, MOS. Mosteller and Tukey [Ref. 3: pp. 315-331] warn that the coefficient of a carrier is very dependent on it's costock. In our case, we will attempt to construct a model so that the carrier variable representing SRB is unrelated to any variable in the costock. The interpretation of the estimated coefficient as *the effect of SRB level*

changing while costock variables keep their same values is then reasonable at the MOS level. For comparisons to be made across different MOS however, we must use the same model for all MOS. While such a solution approach has the disadvantage of suboptimizing our prediction capability at the MOS level, it has the large advantage of permitting a reasonably valid comparison of the relative effectiveness of SRB across a group of MOS.

From the perspective of the user, the overall model approach offers two other distinct advantages. First, it offers simplicity. The managers who will be responsible to implement and maintain this model are not operations analysts and will resist integrating a complicated model / procedure into an already busy schedule. Second, an overall model offers credibility. It would be very difficult to explain to non-analysts why a particular carrier, say Consumer Price Index, is pertinent to the reenlistment decision of a soldier in one MOS, but not in another.

An outline of the steps included in our modelling and solution approach follows. It is consistent with a methodology recommended by Draper and Smith [Ref. 4: p. 414].

1 Define the problem. Select a response variable. Suggest relevant carrier variables.

2 Can we obtain a complete set of observations on all specified carrier variables and the selected response variable? If not, return to step (1). Otherwise, continue.

3 Establish model goals. Consider the minimum / maximum number of included carrier variables desired and determine the desired level of statistical significance for the estimated coefficients of each.

4 Construct a correlation matrix. Guard against including carriers which are highly correlated.

5 Conduct independent multiple linear stepwise regression analysis for each MOS included in the study. Examine the residuals for support of the model assumptions. Are the models adequate? If not, return to step (1). Otherwise, continue.

6 Propose an overall linear regression model.

7 Conduct ordinary least-squares multiple linear regression analysis for each MOS included in the study. Examine the residuals for support of the model assumptions. Is the model adequate? If not, return to step (6). Otherwise, continue.

8 Are the coefficients reasonable? Is the model plausible? Is the equation usable? If not, return to step (1) or (6) as appropriate.

E. INITIAL ASSUMPTIONS

Some further assumptions should be addressed. We assume that an individual's propensity to reenlist is a function of many variables, both personal and environmental. We assume that it is possible to formulate a mathematical model which estimates the propensity of individuals to reenlist at the MOS level. While this assumption is driven by a user requirement for an MOS level model, it is not an unreasonable one. The assumption implies that individuals in the same MOS behave similarly with respect to the factors which affect their reenlistment decision. It also allows that soldiers in different MOS may have different perceptions of the environment in which they make their reenlistment decision. These implications can be justified with respect to the Enlisted Personnel Management System (EPMS). The duties and training required of each MOS are associated with different civilian skills. Also, the general qualifications and skills of the MOS subpopulations are sorted at enlistment. For example, the mean Armed Forces Qualification Test (AFQT) score for one MOS is not the same, nor is it intended to be the same, as any other MOS. EPMS establishes the MOS as the basic unit of personnel inventory management. It is not only the required level, but also the logical level at which to conduct this study.

We must also assume for the purposes of this study that EPMS remains relatively stable. Further, we assume that the socio-economic environment in which the soldier makes a reenlistment decision is stable (within the norms established in the historic scope of this study).

F. THESIS OUTLINE

This thesis formulates and develops a mathematical model which explains the variation in zone A retention at the MOS level. In Chapter II, a brief overview of the SRB program is presented. In Chapter III, the assumptions and analysis leading to the development of an overall model are explained. In Chapter IV, the results of fitting the proposed overall model to the available data are presented and discussed. Finally, Chapter V includes the conclusions and recommendations of this study.

G. PROGRAMMING LANGUAGES AND STATISTICAL PACKAGES.

All programming associated with data collection and manipulation was completed using FORTRAN 77 code. All data analysis and most graphics were completed using the SAS, version V, statistical package. These choices were made with respect to the current capabilities and assets of the Military Personnel Center.

II. THE SELECTIVE REENLISTMENT BONUS PROGRAM

This Chapter presents a brief overview of the Selective Reenlistment Bonus (SRB) program. Criteria for including MOS in the program are outlined, as are the eligibility requirements and payment procedures. Finally, the budget history of the program is graphically summarized.

A. THE OBJECTIVE

The Selective Reenlistment Bonus program is designed to offer an attractive reenlistment incentive to improve manning in critical military specialties.

B. CRITERIA FOR INCLUDING MOS IN THE SRB PROGRAM

As has been previously noted, there are many criteria considered before including, or excluding an MOS from the SRB program. Among these factors are:

1 a comparison of career manning requirements with projected inventory,

2 the cost of formal school training for replacement personnel,

3 the expected increase in retention as a result of inclusion in the SRB program,

4 the priority of MOS in terms of it's *essentiality* to the Army mission,

5 the inherent unattractiveness of the MOS with respect to other military and civilian occupations.

C. ZONES OF ELIGIBILITY

There are three zones of individual SRB eligibility. They are:

1 zone A, which applies to those service members who have completed at least 21 months of continuous active duty but not more than 6 years of active duty on the day of reenlistment.

2 zone B, which applies to those service members who have completed at least 6 but no more than 10 years of active duty on the day of reenlistment.

3 zone C, which applies to those service members who have completed at least 10 but no more than 14 years of active duty on the day of reenlistment.

D. THE AMOUNT OF BONUS AND METHOD OF PAYMENT

1. *Amount of Bonus*

The reenlistment bonus to which a service member is entitled upon reenlistment is computed as follows:

SRB = (monthiy base pay) x (years of additional obligated service) (2.1)

x (SRB level)

where the SRB multiplier can assume values of 0, 1, 2, 3, 4. or 5. No more than one SRB is authorized per soldier per zone. No SRB can exceed $20,000.00.

2. *Method of Payment*

Upon qualification for award of an SRB, a service member receives 50% of the authorized SRB on the day of reenlistment, and the balance in equal annual installments on the anniversary of the reenlistment during the reenlistment contract period.

E. INDIVIDUAL ELIGIBILITY CRITERIA FOR ENLISTED SERVICE MEMBERS.

The individual eligibility criteria for service members is as prescribed in Army Regulation (AR) 600-200 and AR 601-280.

F. PAYMENT EXPERIENCE

As is indicated above, the amount of the SRB award to which an individual is entitled is a function of three factors: SRB award level, individual monthly base pay, and years of additional obligated service incurred as a result of the contract. The two following graphics are included to provide the reader with a feel for the scope of the problem. At Figure 2.1, the horizontal axis lists fiscal years while the vertical axis is scaled to measure the total number of zone A SRB takers for each year. At Figure 2.2, the horizontal axis again represents fiscal years, but the vertical axis represents the total zone A SRB expenditures for each year. We note that both bonus takers and expenditures were at a low point in FY83. We note also that while the total number of zone A bonus takers has increased over the last 2 years, the total expenditures have not. The underlying cause of this trend is that, in general, reenlistment bonuses are available to more eligible soldiers, but at a lower level.

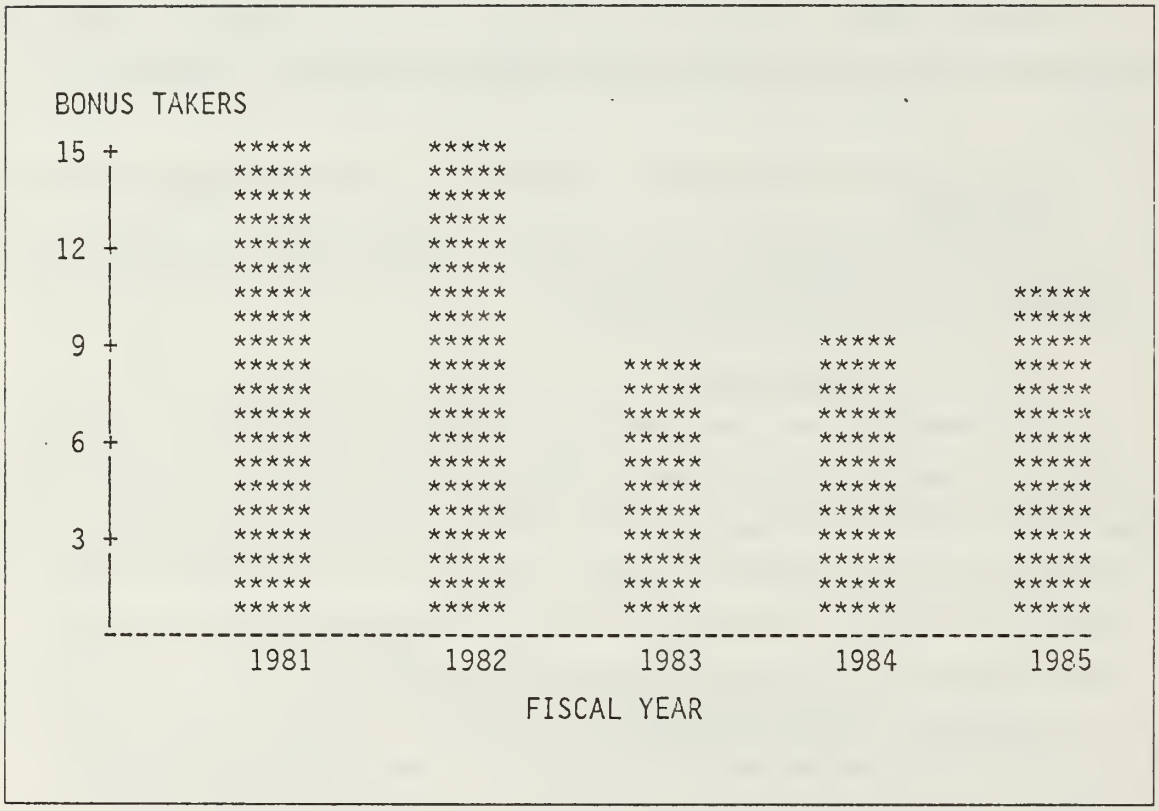


Figure 2.1 Zone A SRB Takers, FY81-FY85
(in thousands)

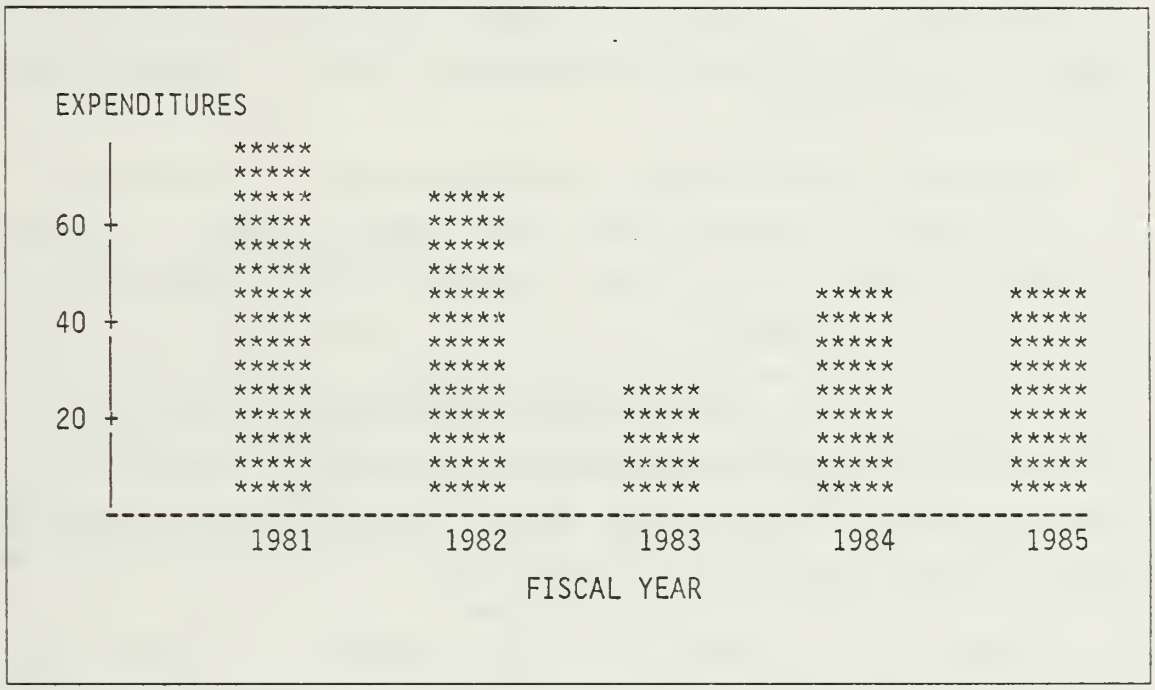


Figure 2.2 Zone A SRB Outlays, FY81-FY85
(in millions of dollars)

III. MODEL FORMULATION

In this Chapter, the assumptions and analysis leading to the development of an overall model are explained. First, the basic multiple linear regression model is proposed in matrix notation. Then a response variable and a set of candidate carrier variables are suggested. A sampling period is defined for use in estimating parameters associated with the proposed variables. The problems encountered in data collection and data preparation are discussed. The results of independent stepwise regression analysis on each of the included MOS are explained. Finally, an overall multiple linear regression model is proposed.

A. PROPOSED LINEAR MODEL IN MATRIX FORM

In this thesis, we assume that there exists a relationship between the propensity of a soldier to reenlist and that soldier's perception of the environment. A reliable method of analysis to examine the nature of the relationship between our proposed response variable (some measure of retention rate) and our candidate carrier variables (which will attempt to account for changes in the makeup or environment of the reenlistment (decision-maker) is the method of least squares, or regression analysis. Using this method of analysis, we will attempt to fit the following multiple linear regression model to the data we collect for each MOS:

$$Y = X\beta + \varepsilon \tag{3.1}$$

where:

Y is an (n x 1) vector of observations on the selected response variable

X is an (n x p) matrix of observations on the selected carrier variables

β is a (p x 1) vector of parameters to be estimated

ε is an (n x 1) vector of errors assumed to have the distribution $N(0, \sigma^2 I)$

It is shown [Ref. 4: pp. 86-87] that if $X'X$ is non-singular, the least squares estimate of β, call it b, can be written as:

$$b = (X'X)^{-1}X'Y \tag{3.2}$$

with variance-covariance matrix $(X'X)^{-1}\sigma^2$. Thus, the variance associated with estimating any particular coefficient is given by:

$$V(b_i) = c_{ii}\sigma^2 \tag{3.3}$$

where c_{ii} is the diagonal element in $(X'X)^{-1}$ corresponding to ith variable. Further, a prediction of Y at X_0 is given by:

$$\hat{Y}_0 = b'X_0 \tag{3.4}$$

with variance given by:

$$V(\hat{Y}_0) = X_0'(X'X)^{-1}X_0(\sigma^2). \tag{3.5}$$

B. SELECTION OF THE RESPONSE VARIABLE

We have assumed that MOS subpopulations can be treated as discrete groups with respect to their propensity to reenlist. Therefore, it follows that if the variables relevant to the reenlistment decision were known, and their levels could be fixed, or considered fixed for a period of time, the reenlistment propensity of these discrete groups could also be considered fixed. Let us assume that these propensities are probabilities. Then, since a soldier either does (1) or does not (0) reenlist, over a period of time we will observe outcomes on repeated bernoulli trials with fixed parameter p.

If we further assume these observations are independent, then we can use the maximum likelihood estimator for parameter p ($\hat{p}$ = number of reenlistments observed / number of trials). Hence, one method for obtaining an estimate of the reenlistment propensity for a given MOS is to observe outcomes on the reenlistment decision for a period of time short enough so that relevant conditions may be fixed or considered fixed, yet long enough to obtain a sample size which will enable us to discern small changes in the population parameter.

The purpose of the SRB program, as stated in Chapter II, is to improve manning in critical military specialties. An SRB can be considered effective in 2 ways. First, an SRB can induce a soldier to reenlist for his own MOS, who may otherwise have left the service. Second, it can induce a soldier to reenlist for his own MOS, who may otherwise have reenlisted for training in another specialty. In conjunction with program managers at MILPERCEN, the following retention (vice reclassification) rate has been developed for use as the response variable in this study:

Y = retention rate = propensity of a soldier to reenlist for his own MOS.

It is estimated by:

$\hat{Y}$ = estimated retention rate = number of soldiers reenlisting for their own MOS / number of soldiers eligible to do so.

Obviously excluded from our estimator $\hat{Y}$ (not included in either numerator or denominator expressions) are service members who are not fully eligible for reenlistment at the decision point. An SRB cannot induce an otherwise ineligible soldier to reenlist. Also excluded are reenlistments which occur outside the window of eligibility (6 months for first term soldiers, 3 months otherwise) and all extensions. These actions, while not independent of the effects of the SRB program, occur for exceptional reasons unrelated to the SRB award level. Soldiers who reenlist, but reclassify in conjunction with reenlistment, are not counted in the numerator of our estimator, but are included in the denominator.

Retention data is available at the individual soldier level on mass storage at MILPERCEN. However, owing to significant changes in the manner in which these data were recorded prior to fiscal year 1981, earlier data are not readily available. A magnetic tape, containing information pertinent to the reenlistment or separation of soldiers during the period 1 Oct 81 through 30 Sep 85, was provided by MILPERCEN to support this study. Excluded from this tape were transactions concerning service members outside of the three SRB zones, or who otherwise fell into an excluded category as described in the previous paragraph. In all, more than 481,000 individual records were included in the file.

C. SELECTION OF THE CARRIER VARIABLES

1. *SRB Level*

SRB level is the carrier variable of interest in this study. It exists at one of 6 discrete levels for all MOS, for all zones, at all times. These levels are 0, 1, 2, 3, 4, and 5. Record of the SRB history for each MOS is not currently available in machine readable form, but hardcopy records were made available by the MILPERCEN program managers dating back to 1974.

2. *Endogenous Variables*

The endogenous variables, for the purposes of this study, are those variables which provide information on the demographic composition of the discrete groups themselves. For each record contained on the data tape provided by MILPERCEN, the following demographic data are recorded:

1 AFQT score,

2 civilian education level,

3 sex,

4 number of dependents,

5 race.

It is our intention in recording these data, to construct variables which may be included in the overall regression model to control for the effects of population dynamics.

3. *Exogenous Variables*

Unemployment rate is included as a statistic which is visible to the reenlistment decision-maker and may represent one quantitative measure of the soldier's career alternatives. This data is readily available in the *Employment and Earnings Monthly*, published by the Bureau of Labor Statistics (BLS). The data is summarized by occupational classification and region. Since most Army skills do not readily fall into any of the BLS classifications, our statistic of choice is the seasonalized aggregate unemployment rate.

Consumer Price Index (CPI), as a measure of the change in the spending power of the soldier, is also considered a vital statistic. Data is again available on a monthly basis in the BLS published *CPI Detailed Report*. The statistic most relevant for our uses is the seasonalized statistic for all urban consumers.

Pay scale changes are believed to be at least as important as CPI. Considered with CPI, a measure of the real change in a soldier's purchasing power can be derived.

Promotion opportunity to pay grades E5 and E6 is considered very important. Variables which account for the change in promotion opportunity at the MOS level were of specific concern to the MILPERCEN program managers. Our problem here however, is to identify a measure visible to the reenlistment decision-maker and for which a reliable historic record exists. The monthly published promotion cut-off scores were an immediate choice as an explicit and simple indicator of relative promotion opportunity, but MILPERCEN promotion program managers have maintained no data older than 2 years. As an alternative, it was decided to include a statistic reported on the monthly DCSPER 411, Enlisted Strength Report, available on microfiche only. The statistic, *mean time in service at promotion for those promoted in the previous 12 months*, reports a 12 month promotion moving point average for both grades at the MOS level. This statistic is included, as it is believed that a soldier making a reenlistment decision is sensitive to the effects changes in promotion policy have on the careers of those around him.

D. SELECTION OF A SAMPLE PERIOD

As has been mentioned, our data collection capability is limited to the five fiscal years from FY81 through FY85. A change in the manner in which loss data was recorded precludes our obtaining reliable data on earlier records.

Inasmuch as we plan to observe outcomes on the reenlistment decision over a period of time during which the levels of the independent variables included in our regression model are considered fixed, we must decide upon a sample period. An immediately attractive alternative is the fiscal quarter for several reasons. First, the SRB program is managed in accordance with a quarterly cycle. Second, several of our data (such as the promotion statistics) are reported at quarterly intervals. Third, several of our data (such as CPI) are much more stable at the quarter level.

Analysis was conducted to determine the appropriate sample size of eligibles required to ensure that a reliable base of MOS and zone specific retention rate estimates was obtained. Specifically, we wish our sample size to be large enough so that 90% of the time our estimate $\hat{Y}$ is within 10% of the true parameter Y. Then using an approximate 90% confidence interval for for the Bernoulli parameter Y [Ref. 5: pp. 394-395], we can compute the minimum number of observations, n, required to satisfy our requirement. The approximate 90% confidence interval can be written as:

$$P(\hat{Y} - 1.645(\hat{Y}\,(1-\hat{Y})/n)^{1/2} < Y < \hat{Y} + 1.645(\hat{Y}\,(1-\hat{Y})/n)^{1/2}) = .90$$

The variance of the estimate is *maximized* with $\hat{Y} = 0.5$.

$$P(.5 - 1.645(.25/n)^{1/2} < Y < .5 + 1.645(.25/n)^{1/2}) = .90$$

We see that to be 90% confident that our estimate $\hat{Y}$ is within 10% of the true parameter Y, it must be true that:

$$1.645(.25/n)^{1/2} < .10$$

Solving the above equation for n, we find that:

$$n > 68$$

We next require each MOS included in our analysis to have at least 68 zone A reenlistment outcomes per quarter for no fewer than 14 of the 20 quarters of data available. We will refer to such MOS as high density. In addition we require that the MOS be authorized as of the end of FY85 and that it have an active SRB history in our period of study. That is, there must be at least one change in SRB level during the data period. When these requirements are imposed, the number of MOS included in our analysis is reduced from an initial 374 to 24. These MOS are listed in Table 1.

Consider the SRB budget history summarized at Figure 2.2. While the number of MOS included in our analysis represents only 6.4% of the total MOS in the inventory, during the 5 year period of our study, these 24 MOS accounted for over 34% of the zone A reenlistments and over 60% of the total zone A bonus budget outlays. With these facts in mind, we will pursue our development of a zone A retention model using only the 24 high density MOS. In doing so, we make the following observations:

1 The developed model should be accurate for the 24 high density MOS.

2 Inasmuch as the model will account for over 34% of the total zone A reenlistments in the Army, it is very likely to be reasonably accurate for the moderate density MOS in the inventory. (A moderately dense MOS is one for which at least 17, but less than 68, outcomes per quarter can be observed for no fewer than 14 of the 20 quarters of data available for our study. The requirement for 17 observations allows us 90% confidence that our estimator is within 20% of the true retention rate, Y.) An application of the developed model to those MOS will not be unjustified.

3 It may not be possible to adequately represent the retention behavior of all low density MOS with an overall model. By their nature, they are managed exceptionally. Their group perception of the factors which affect their reenlistment decision will not likely be similar to that of any other MOS group. Efforts to group these low density MOS, creating artificial high density sample cells, as has been done in several studies by both CAA and Rand Corporation (including those previously referenced), must be well documented and controlled.

TABLE 1

MOS INCLUDED IN THIS ANALYSIS
(HIGH DENSITY)

MOS	TITLE
11B	Infantryman
11C	Indirect Fire Infantryman
11H	Heavy Anti-armor Weapon Infantryman
12B	Combat Engineer
12C	Bridge Crewman
13B	Cannon Crewmember
13E	Cannon Fire Direction Control Specialist
13F	Fire Support Specialist
16R	ADA Short Range Gunnery Crew Member
16S	MANPADS Crewmember
19D	Cavalry Scout
19E	M48-M60 Armor Crewmember
31M	Multichannel Commo Equip Operator
31V	Tactical Commo Equip Operator
51B	Carpentry / Masonry Specialist
54E	NBC Specialist
63B	Light Wheel Vehicle Mechanic
63H	Track Vehicle Repairer
63N	M60A1/A3 Tank System Mechanic
63T	Bradley FVS Mechanic
63W	Wheel Vehicle Repairer
72G	Telecommunications Center Operator
76W	Petroleum Supply Specialist
82C	Field Artillery Surveyor

It is acknowledged here that our approach to the sample size problem is very conservative. We will show in Chapter IV, that actual results from applying our proposed linear model to available data for high density MOS, can yield 90% confidence intervals which are considerably shorter than (+/-)10%.

E. DATA PREPARATION

The zone A SRB level in effect for each MOS and for each quarter is included in the candidate carrier variable data set (as variable SRB) without modification. An additional variable, SRBSQ (SRB^2) is also included to account for the possible nonlinear effects of the SRB program on retention.

The FORTRAN code which was used to develop retention rates (response variable REUP) and other rates associated with the endogenous variable set is included at Appendix A. The retention rate algorithm is straightforward and consistent with the rules set forth in section B of this Chapter. The endogenous carrier variables are defined for each of the 24 MOS and for each of the 20 quarters as follows:

1 AFQT : eligible population scoring less than 50 on the AFQT / total eligible,

2 CIVED : eligible population completing at least 12 years of formal education / total eligible,

3 SEX : eligible females / total eligible,

4 DEP : eligible population with more than 2 dependents / total eligibles,

5 RACE : eligible non-caucasians / total eligible.

Initial demographic rate definitions were suggested by retention program managers at MILPERCEN. The final definitions reported above were developed through a trial and error process. These definitions were found to provide the most meaningful description of an eligible population.

A variable named REAL was constructed as a linear combination of the CPI and the annual pay raise received by the service member. Specifically, REAL = % pay raise - CPI. The variable was considered as a carrier because we found that it adequately accounted for the changes in the soldier's purchasing power, while consuming one fewer model degrees of freedom.

The E5 and E6 promotion opportunity variables included in the candidate carrier variable set were constructed as follows:

1 E5TEST2 : mean time in service (TIS) at promotion to grade E5 for those promoted in the previous 12 months (MOS level) / mean TIS at promotion to grade E5 for those promoted in the previous 12 months (Army level).

2 E6TEST2 : mean TIS at promotion to grade E6 for those promoted in the previous 12 months (MOS level) / mean TIS at promotion to grade E6 for those promoted in the previous 12 months (Army level).

We expect to find that E5 and E6 promotion opportunity (here, measured relative to an Army average) are effective retention incentives. That is, as the relative opportunity for promotion in a particular MOS is enhanced, so should the retention rate be enhanced, given the levels of all other factors are unchanged.

The seasonally adjusted unemployment rate (UNEMPLY) is included in the candidate variable set without modification.

Our earliest analysis of the data provided by MILPERCEN indicates the existance of a strong seasonal trend in retention. Figure 3.1 graphically depicts this trend. The solid line represents the aggregate estimated retention rate for all MOS which were *not* included in the SRB program during our period of analysis. The broken line represents the aggregate estimated retention rate for all MOS which were included in the SRB program during our period of analysis.

Three observations can immediately be made. First, the aggregate trends are very similar. Second, despite the inducement of a bonus, MOS included in the SRB program tend to have lower rates of retention than those not included. Third, and most importantly, it is evident that we could capture a good deal of the seasonality by including the variables QTR (representing the actual fiscal quarter associated with each data point and taking on values 1, 2, 3 or 4) and QTRSQ (QTR^2) in the candidate variable data set. A variable or set of variables which accurately accounts for an effect such as seasonality is preferred to an explicit representation of the cause when, such as in our case, the result is a large reduction in model degrees of freedom.

F. THE STEPWISE REGRESSION MODEL

Stepwise regression is a method of building a multiple linear regression model using only the best independent carrier variables. In stepwise regression, we first construct a first order linear regression model using only that independent variable which is most highly correlated with the designated response variable. We check the results of an overall F-test to determine if our regression is significant at some pre-selected level. If not, we discontinue our analysis and select $\hat{Y} = \overline{Y}$ as our best predictor. Otherwise, we retain that initial variable in our model and search for a second significant carrier variable to enter the regression. The partial correlations of each of the remaining candidate carrier variables with the response variable are examined and the variable with the highest partial correlation is added to the regression. The partial F-statistics of each carrier variable included in the model are

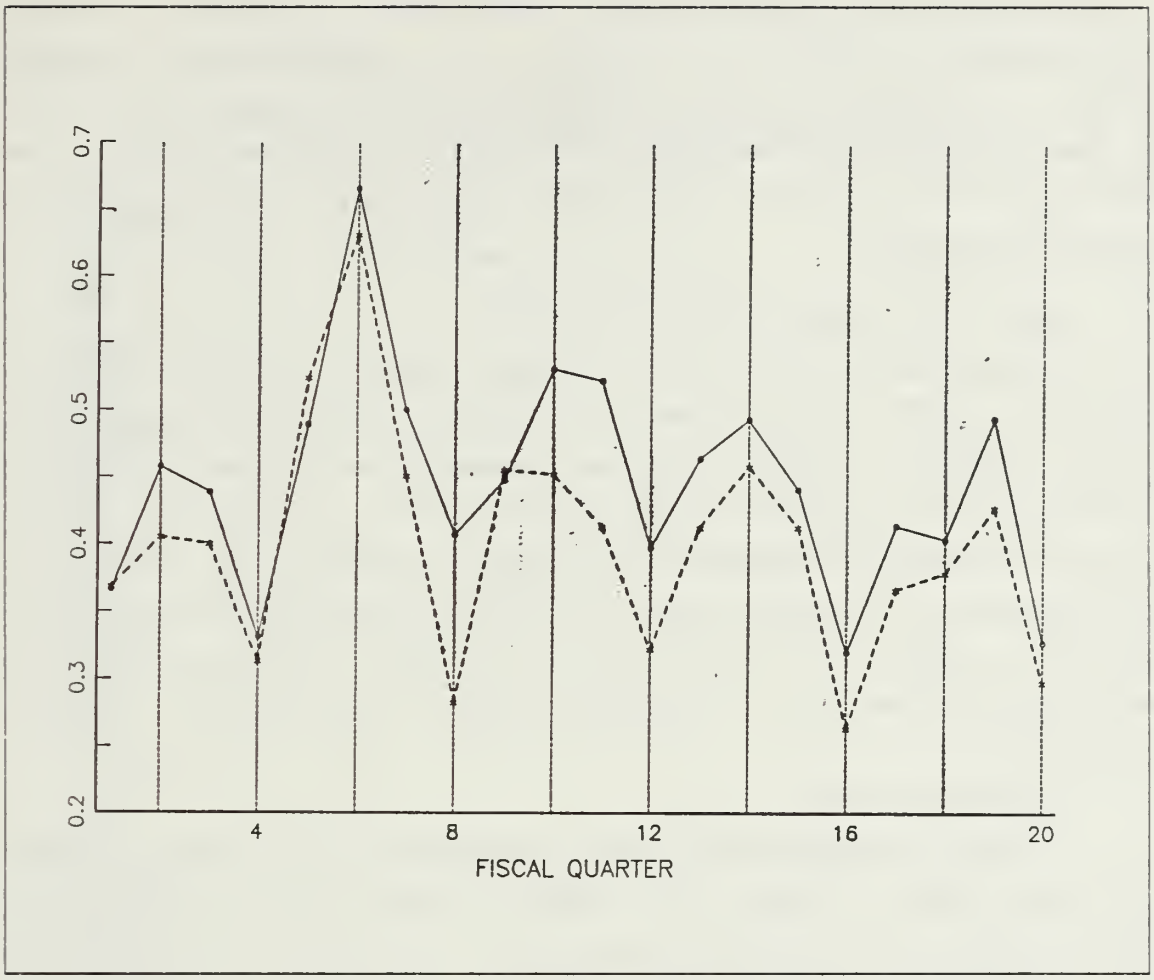


Figure 3.1 The Seasonality of Retention
(Bonus and Non-bonus MOS)

examined and compared to a pre-selected acceptance level. If they are both significant, they are retained and a third candidate carrier is proposed. Otherwise, we eliminate the non-significant carrier(s) from the regression model and identify the next best candidate. This process is continued until the set of variables included in the model cannot be altered at the pre-selected significance level. [Ref. 4: pp. 306, 312].

The correlation matrix of the response variable and each of the candidate carrier variables for all data in our data set (24 MOS x 20 observations per MOS = 480 observations) is at Appendix B. Note that the variable SRB is more highly correlated with the response variable REUP than any other. There do not appear to be any dangerous correlations among the candidate carriers at the aggregate level. Recall, we wish to guard against any singularity or near singularity of the X'X matrix.

An example of an input data set for MOS 63B is at Appendix C. Note that variables SRBSQ and QTRSQ do not appear, as they are constructed in the modelling process. An example of the output from a SAS STEPWISE procedure is at Appendix D. Precise instructions for interpreting this output are contained in [Ref. 6] and [Ref. 7: pp. 761-774]. The SAS commands which were used to generate this output are included in Appendix G.

G. RESULTS OF THE STEPWISE ANALYSIS

We summarize the results of our stepwise analysis in three ways. First, we examine the results of each regression to determine which carrier variables had estimated regression coefficients which were reasonably and consistently signed and significant at the .15 acceptance level most often. Then, as a measure of the total variation in retention rate explained by our model, we examine the R^2 statistic for all MOS included in our analysis. Finally as a measure of goodness of fit, we examine Mallows C_p statistic for all MOS included in our analysis. After we have proposed and applied an overall model, a more detailed analysis of model residuals is presented in Chapter IV.

1. *Significant Carriers*

In Table 2, each candidate carrier variable is listed. The pair SRB* / SRBSQ* and the pair QTR* / QTRSQ* are also included and will be used to record the event that *both* carriers were considered significant for a particular MOS. For example, if SRB and SRBSQ are both included for some MOS, an observation will not be recorded for the carriers SRB and SRBSQ. Instead an observation will be recorded for both SRB* and SRBSQ*. Observations for SRB and SRBSQ (or QTR and QTRSQ) are recorded only when they are un-paired. An observation for any candidate variable is recorded when the variable has been included included in the stepwise model at the .15 level of significance. The manner of record chosen (+ / -) indicates the sign of the estimated coefficient.

We note in Table 2 that the SRB* / SRBSQ* pair is not often significant while the QTR* / QTRSQ* pair is. However, we also note that the variables SRB or SRBSQ, or their pair, are considered significant in 17 of the 24 individual models examined. Other variables which appear to be excellent carrier candidates are RACE, DEP and REAL.

TABLE 2

SIGNIFICANCE OF CARRIERS (STEPWISE PROCEDURE)
(0.15 SIGNIFICANCE LEVEL)

CARRIER	RESULTS
SRB	+ + + + + + + + + + +
SRBSQ	+ + + + +
SRB*	+
SRBSQ*	-
QTR	-
QTRSQ	- - -
QTR*	+ + + + - + + + +
QTRSQ*	- - - - + - - - -
RACE	+ + + + + + + +
DEP	+ - - + + + + + + + + +
EDUCATE	- - + - + - + -
AFQT	+ + + - -
E5TEST2	+ + + + - -
E6TEST2	+ + + -
UNEMPLY	+ + + +
REAL	+ + + - + + + + + + +

2. *The R^2 Statistic*

A commonly accepted statistic for measuring the value of a regression equation is the R^2 statistic. The R^2 statistic actually measures the proportion of total variation about the mean, $\overline{Y}$, which is accounted for by the regression. We are cautious in using this statistic, because it can be made arbitrarily high by adding different, albeit meaningless carriers [Ref. 4: p. 33].

With this caution in mind, the results of our R^2 analysis are summarized in Figure 3.2. The horizontal axis is grouped into R^2 bins of width 0.1, while the vertical axis represents the number of occurances.

3. *The Mallows C_p Statistic*

Another popular statistic for measuring the goodness of fit for a proposed model is the C_p statistic developed by C. L. Mallows [Ref. 4: pp. 299, 303]. The expected value of the statistic is approximately the number of independent carriers included in the regression model plus the intercept term (p). Extraordinarily high values of the C_p statistic indicate that our model suffers considerably from lack of fit; that is, our residuals are composed of both random and systematic components. In our analysis of the given data, we find that three of the proposed regression models obtained via the stepwise procedure suffer from lack of fit. They are the models associated with the MOS listed in Table 3. We will pay particular attention to these MOS in attempting to fit an overall model.

II. THE PROPOSED OVERALL MODEL

The proposed overall model, based on the requirements of the study and the previous analysis, is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5 + \varepsilon \qquad (3.6)$$

where:

Y = retention rate (as previously defined)
X_1 = SRB
X_2 = QTR
X_3 = RACE
X_4 = DEP
X_5 = REAL
ε = error component with assumed distribution $N(0, \sigma^2)$

and β is a vector of the parameters to be estimated.

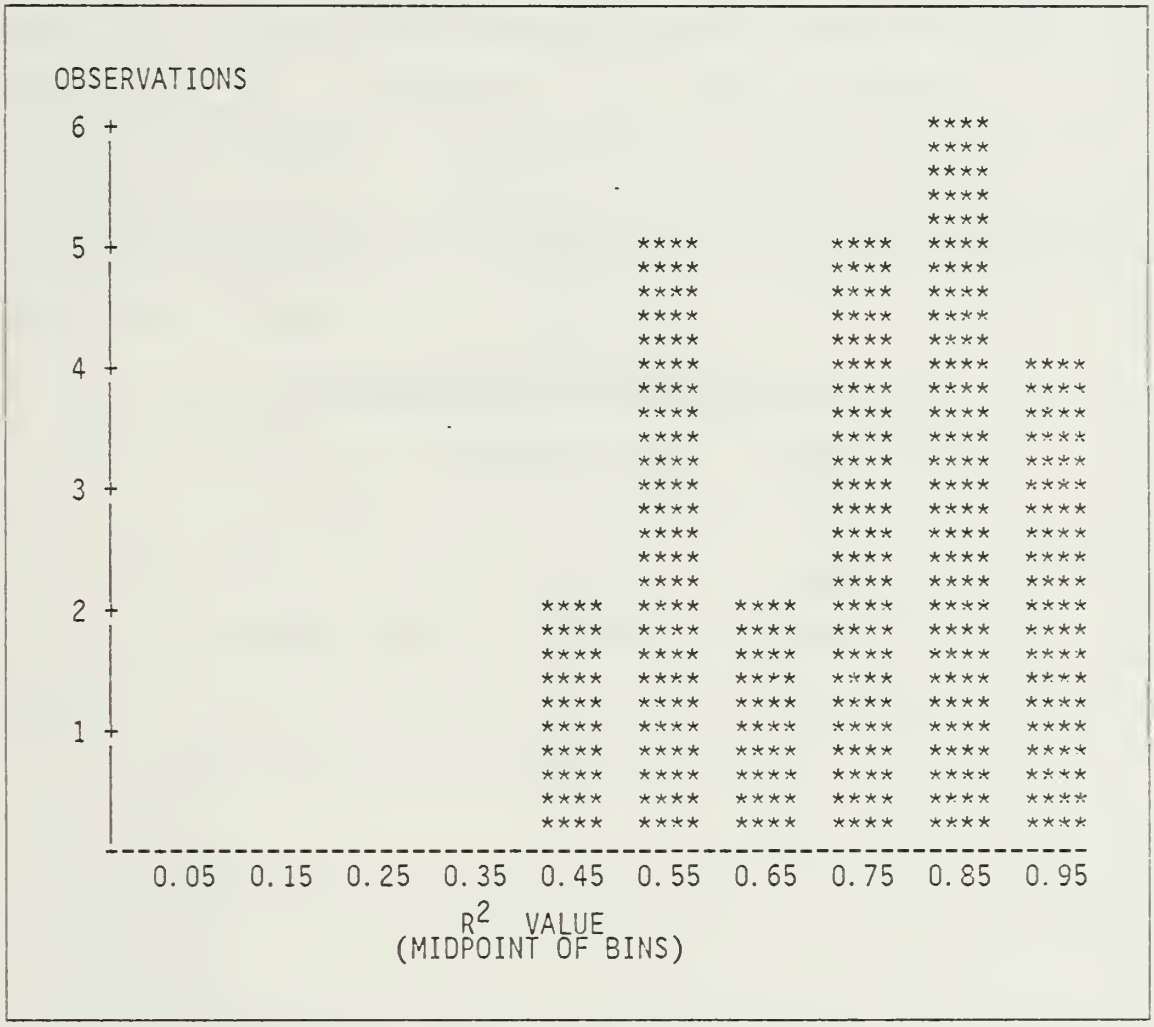


Figure 3.2 Distribution of R^2 Values
(Stepwise Procedure)

TABLE 3

LACK OF FIT MODELS
(FROM THE STEPWISE PROCEDURE)

MOS	C_p Statistic	p
12B	47.25	3
31M	36.76	4
51B	35.21	3

IV. THE ZONE A RETENTION MODEL

In this Chapter, ordinary least squares multiple linear regression analysis is used to fit the overall model proposed in Chapter III to the data available for the high density MOS. The results of this analysis are discussed in terms of carrier significance and the R^2 statistic. An examination of the residuals is performed to investigate suspected model inadequacies. The model is then fit to data available for the moderate density MOS. The results of this analysis are briefly summarized and potential data transformations are discussed. A demonstration of the uses of this model in both a predictive and comparative mode is presented. Finally, alternatives for modelling low density MOS are suggested.

A. THE OVERALL MODEL FITTED TO HIGH DENSITY MOS

The overall model, as proposed in the previous Chapter, is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 \qquad (4.1)$$

$$+ \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5 + \varepsilon$$

where:

Y = retention rate (as previously defined)
X_1 = SRB
X_2 = QTR
X_3 = RACE
X_4 = DEP
X_5 = REAL
ε = error component with assumed distribution $N(0, \sigma^2)$

and β is a vector of the parameters to be estimated.

In applying ordinary least squares linear regression analysis to our data, we recognize that we have 20 unadjusted degrees of freedom (df) available for each MOS (via our 20 quarterly observations on the response and carrier variables). Our proposed model requires 1 df for the intercept estimate, b_0, and 6 df for the proposed carrier variables, leaving 13 df for error. While no hard and fast rules exist for the

optimal distribution of available df in the development of a linear model, a good rule is to keep the model degrees of freedom (in our case, 7) small relative to the total available degrees of freedom. This is a particularly good rule when the model degrees of freedom are limited, as they are in our analysis.

The proposed overall model was fitted to the data available for the 24 high density MOS. The SAS commands which were used to generate our output are included at Appendix G. A copy of our output for example MOS 63B is at Appendix E.

We can easily summarize our results of this analysis in a manner similar to that used for our stepwise analysis in the previous Chapter. First, we examine the estimated coefficients of each carrier for each MOS to determine which were most often consistent and most often significant. We note that our results for the included carriers may well differ from the results we obtained for those same carriers in our stepwise procedure. Despite our efforts to select candidate carriers which were unrelated, it is very possible that for a particular MOS, a carrier which was included (AFQT, for example) in the stepwise model served as a *proxy* [Ref. 3: p. 317] for some carrier which was not included (say, DEP). Since DEP is included in the overall model, and AFQT is not, it would not be surprising if DEP were to suddenly *become* significant at the .15 level in our current analysis, even though it was rejected at that same level in our stepwise analysis. This phenomenon is a consequence of our resolve to develop an overall model.

After our estimated coefficient analysis, we will present an R^2 statistic summary, similar to that presented in Chapter III.

1. *Significant Carriers*

In Table 4, a summary of the results in terms of significant carriers using ordinary least squares multiple regression analysis is presented. The same definitions for QTR* and QTRSQ* apply as in Chapter III; that is, they represent paired observations on the variables QTR and QTRSQ. We notice that our results from this analysis are very similar to those summarized at Table 2 for the stepwise analysis for all variables except REAL. Previously, REAL was significant at the .15 acceptance level a total of 10 times. In our current analysis, it is significant 18 times, or as many times as the variable SRB is significant.

At the individual MOS level, we can compare our output for MOS 63B via the stepwise procedure (Appendix D) to the output generated when the overall model was

TABLE 4

SIGNIFICANCE OF CARRIERS (REGRESSION PROCEDURE)
(0.15 SIGNIFICANCE LEVEL)

CARRIER	RESULTS
SRB	+ + + + + + + + + + + + + + + + + +
QTR	-
QTRSQ	- - - - -
QTR*	+ + + + + + + + + + + +
QTRSQ*	- - - - - - - - - - - -
RACE	+ + + + + + + + +
DEP	+ + + + + + +
REAL	+ + + + + + + + + + + + + + + + + +

fitted (Appendix E). We note that the carriers which were considered significant via the stepwise analysis, and which were also included in the overall model, remain significant. Carrier variables DEP and REAL, which were not considered significant via the stepwise procedure, are also not significant at the .15 level in our current analysis, although their estimated regression coefficients are signed as expected. The general effect of using an overall model, vice an MOS specific model, in this case is not great. The R^2 statistic has been reduced from .93 to .87, and the overall significance level of the regression has been slightly increased, owing to a slightly larger *error mean square* value.

Note that a critical point made earlier in this thesis is supported by our current analysis. The estimate of an individual regression coefficient is dependent, in varying degrees, on it's costock. The estimate b_1, with costock including E5TEST2 and UNEMPLY (via the stepwise procedure) is valued at .255. With E5TEST2 and UNEMPLY removed, and with DEP and REAL included, the estimate b_1 is increased to .304. While this difference may seem slight (and *is* with respect to the standard error of the estimate), it could be a very significant difference if this coefficient is used as a point estimate of the effectiveness factor (as discussed earlier).

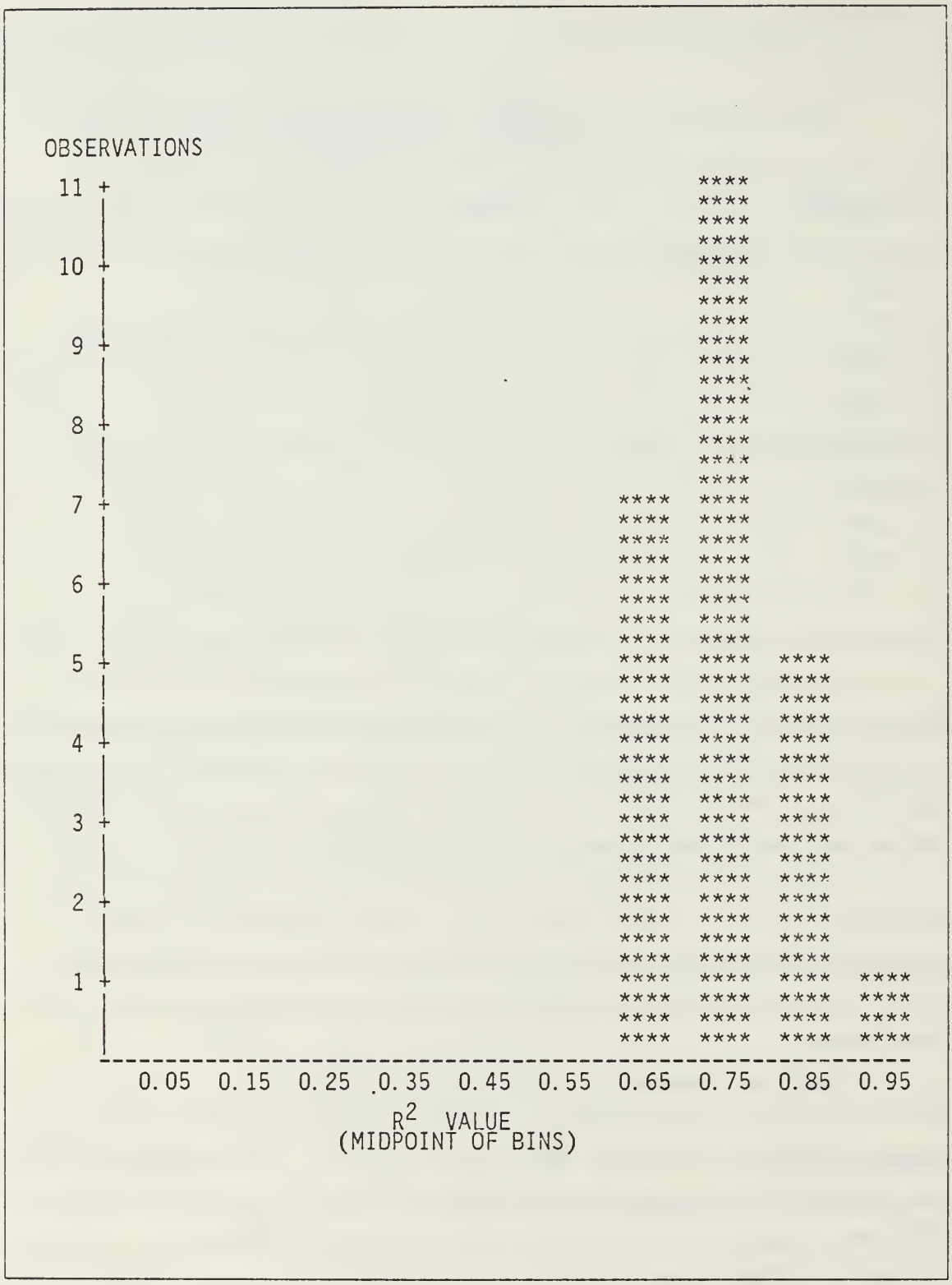


Figure 4.1 Distribution of R^2 Values
(Regression Procedure - High Density MOS)

2. *The R^2 Statistic*

In Figure 4.1, we note that our lowest observed R^2 value is in the .65 bin, whereas in our stepwise summary at Figure 3.2, it was in the .45 bin (an improvement in the distribution of the R^2 values). We note also that the number of observations in the .95 bin has been reduced from 4 in Figure 3.2 to 1 in Figure 4.1. We have examined a case in the previous section wherein the R^2 value moved from the .95 bin to the .85 bin (MOS 63B). MOS 16R is an example of an MOS which moved from the .45 bin to the .65 bin in our analysis.

The actual difference in R^2 values for MOS 16R is .60 - .48 = .12. In the stepwise procedure, only DEP and QTRSQ were included as significant carriers (at the .15 level of acceptance). When the overall model was fitted to the data available for MOS 16R, the other 4 carrier variables were not significant at the .15 acceptance level, but all were signed as we expect, and some variables, such as SRB, were significant at only slightly higher levels. In all, while the R^2 statistic was increased for this MOS, and the sum of squares due to regression was increased, the overall significance of the regression was slightly reduced by the inclusion of the *non-significant* carrier terms.

B. EXAMINATION OF RESIDUALS

Our residual analysis associated with fitting the proposed overall model to the data available for the 24 high density MOS is summarized in the 4 graphics below. The residuals of the 24 MOS were examined independently during the analysis phase of this study, but are here presented in an aggregate manner with enhanced effect.

In conducting a residual analysis, we are examining the validity of the model assumptions concerning the observed errors; that is, that they are independent, have a 0 mean, have a constant variance, and follow a normal distribution. At the conclusion of our analysis, we should observe that either our model assumptions appear to be violated or they do not appear so. [Ref. 4: pp. 141-142].

1. *The Frequency Plot*

In Figure 4.2, we present a horizontal bar chart of the residuals, from -.3 to + .3 in bins of width .01. The distribution of these residuals should appear symmetric (specifically, bell shaped), and centered on 0. No contradiction to our normality assumption is evident here.

2. *The Plot against Fitted Values*

In Figure 4.3, we present a plot of the residuals verses the fitted values associated with them. We hope to find no regular pattern in the residuals; that is, if our model assumptions are correct, the distribution of the residuals is independent of the fitted values. No contradiction to this assumption is evident.

3. *The Plot against Time Sequence*

As in the plot against the fitted values, we should observe no patterns of significance in the plot of residuals verses sequence of observation. In Figure 4.4, while we note a tendency for positive valued residuals associated with observations 6 and 9, they are not abnormally low or high and no regular patterns are discernable.

4. *The Serial Correlation Plots*

In Figures 4.5 and 4.6, we test for Lag-1 and Lag-4 serial correlation respectively. If our observed errors are pairwise uncorrelated, then a *cloud* centered on coordinate (0, 0) should be the only discernable pattern. The Lag-4 plot is suggested by our suspicion that some seasonality effects remain, even after the addition of the QTR and QTRSQ variables to our overall model. It is seen that our suspicions are unfounded.

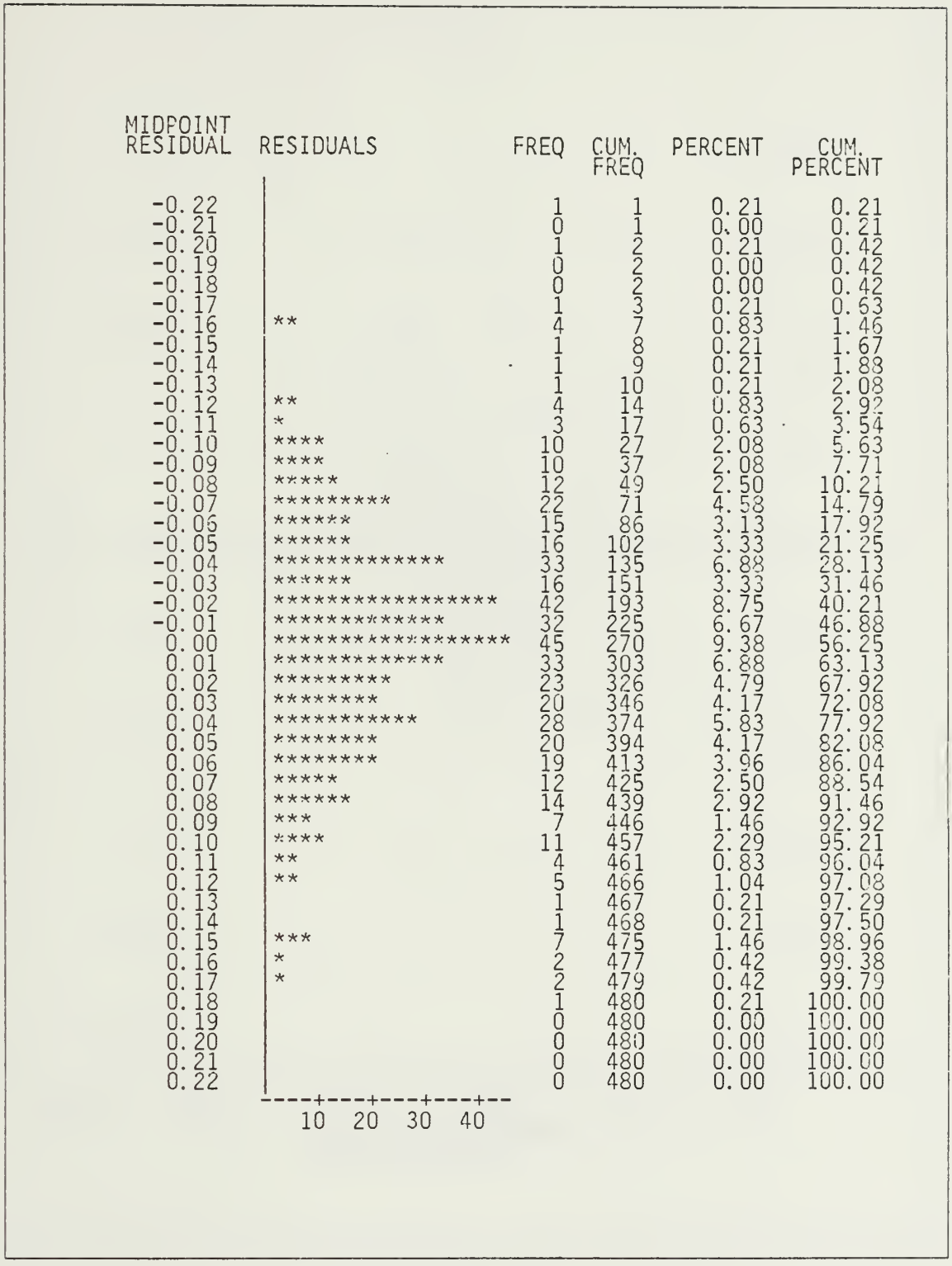
With these results in hand, we are prepared to accept our modelling assumptions as reasonable. The SAS commands which were used to produce all the previous residual graphics are included at Appendix G.

C. THE OVERALL MODEL FITTED TO MODERATE DENSITY MOS

We now have an opportunity to verify our proposed overall model with a fresh data set. From among the remaining MOS, we selected 50 moderate density MOS for which we have record of an active SRB history during the fiscal years 1981-1985. Data for these MOS were gathered in the same manner as for the 24 high density MOS. The proposed linear model was fitted to these data and the results from the 50 independent fittings are summarized, in the aggregate, as follows:

1. *Significant Carriers*

The primary carrier variable of interest, SRB, continues to serve as an excellent predictor variable. In our current analysis, it is significant at the .15 acceptance level in 27 of the 50 moderate density models. The pair QTR and QTRSQ were also included as significant in 27 of 50 cases. The carriers RACE, REAL and DEP were not considered to be as significant as often (14, 14 and 11 times



MIDPOINT RESIDUAL	RESIDUALS	FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
-0.22		1	1	0.21	0.21
-0.21		0	1	0.00	0.21
-0.20		1	2	0.21	0.42
-0.19		0	2	0.00	0.42
-0.18		0	2	0.00	0.42
-0.17		1	3	0.21	0.63
-0.16	**	4	7	0.83	1.46
-0.15		1	8	0.21	1.67
-0.14		1	9	0.21	1.88
-0.13		1	10	0.21	2.08
-0.12	**	4	14	0.83	2.92
-0.11	*	3	17	0.63	3.54
-0.10	****	10	27	2.08	5.63
-0.09	****	10	37	2.08	7.71
-0.08	*****	12	49	2.50	10.21
-0.07	*********	22	71	4.58	14.79
-0.06	******	15	86	3.13	17.92
-0.05	******	16	102	3.33	21.25
-0.04	**********	33	135	6.88	28.13
-0.03	******	16	151	3.33	31.46
-0.02	**********	42	193	8.75	40.21
-0.01	**********	32	225	6.67	46.88
0.00	**********	45	270	9.38	56.25
0.01	**********	33	303	6.88	63.13
0.02	*********	23	326	4.79	67.92
0.03	********	20	346	4.17	72.08
0.04	**********	28	374	5.83	77.92
0.05	********	20	394	4.17	82.08
0.06	********	19	413	3.96	86.04
0.07	*****	12	425	2.50	88.54
0.08	******	14	439	2.92	91.46
0.09	***	7	446	1.46	92.92
0.10	****	11	457	2.29	95.21
0.11	**	4	461	0.83	96.04
0.12	**	5	466	1.04	97.08
0.13		1	467	0.21	97.29
0.14		1	468	0.21	97.50
0.15	***	7	475	1.46	98.96
0.16	*	2	477	0.42	99.38
0.17	*	2	479	0.42	99.79
0.18		1	480	0.21	100.00
0.19		0	480	0.00	100.00
0.20		0	480	0.00	100.00
0.21		0	480	0.00	100.00
0.22		0	480	0.00	100.00

10 20 30 40

Figure 4.2 Residual Bar Graph

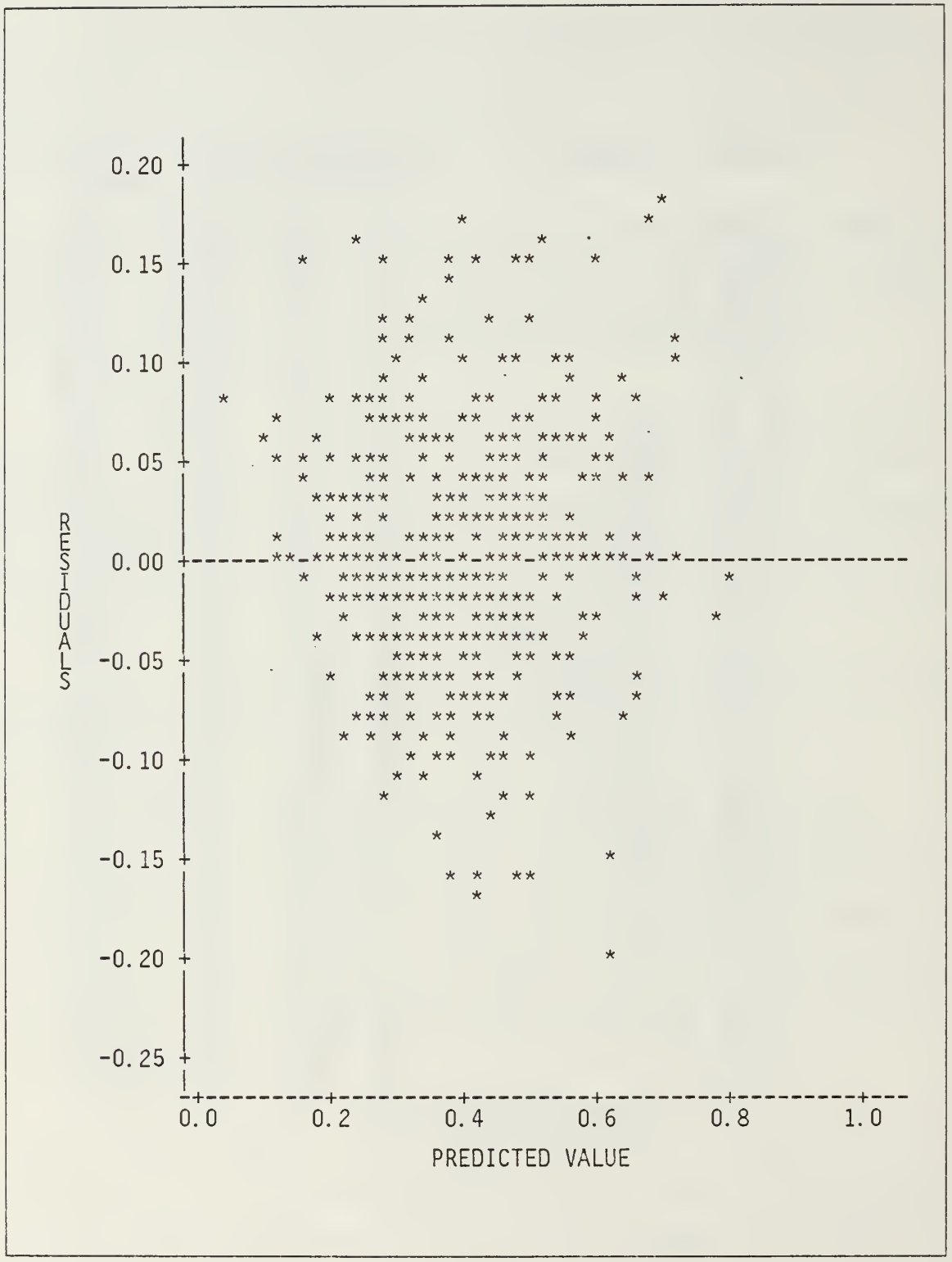


Figure 4.3 Residuals vs. Fitted Values

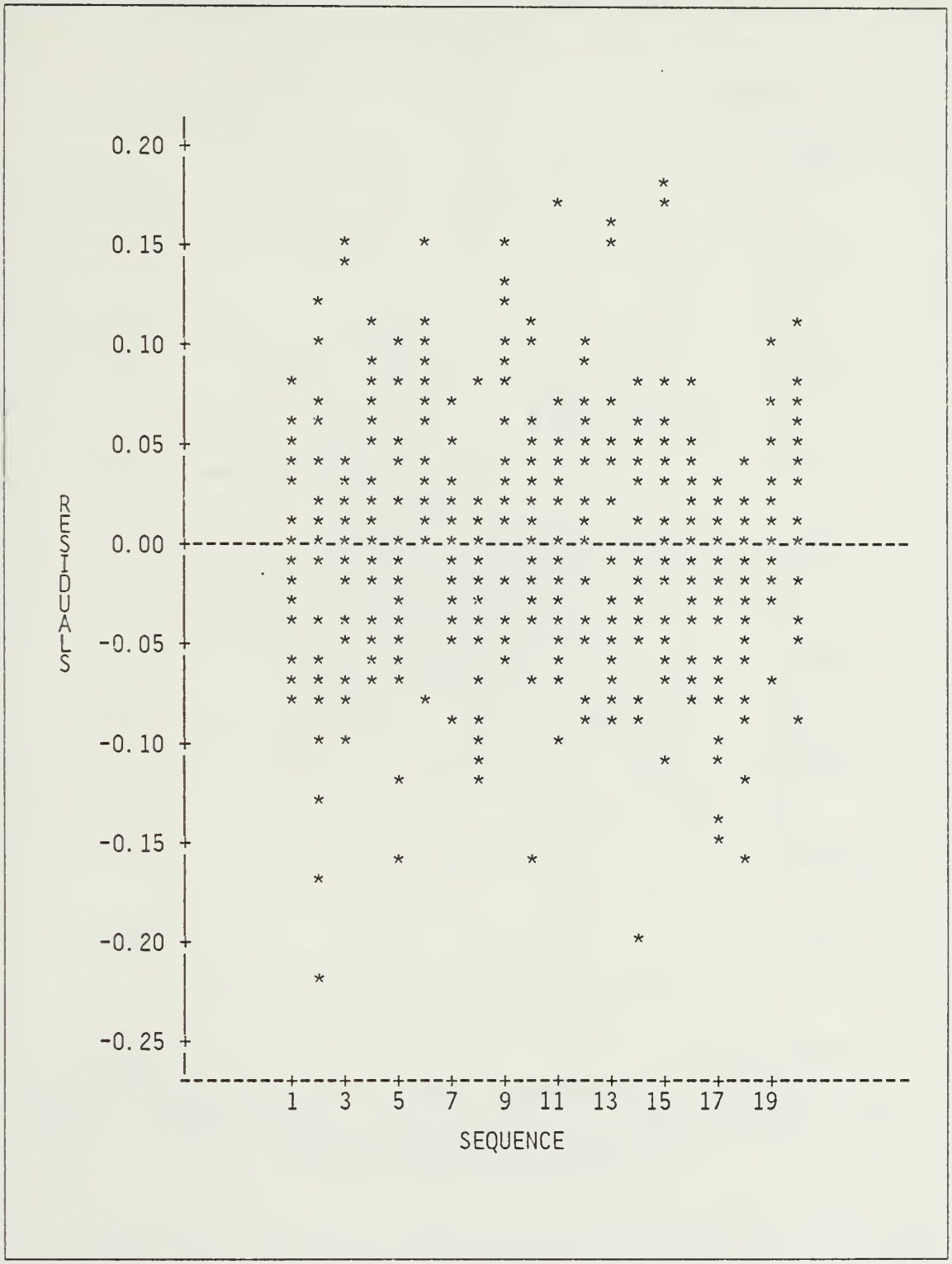


Figure 4.4 Residuals vs. Time Sequence

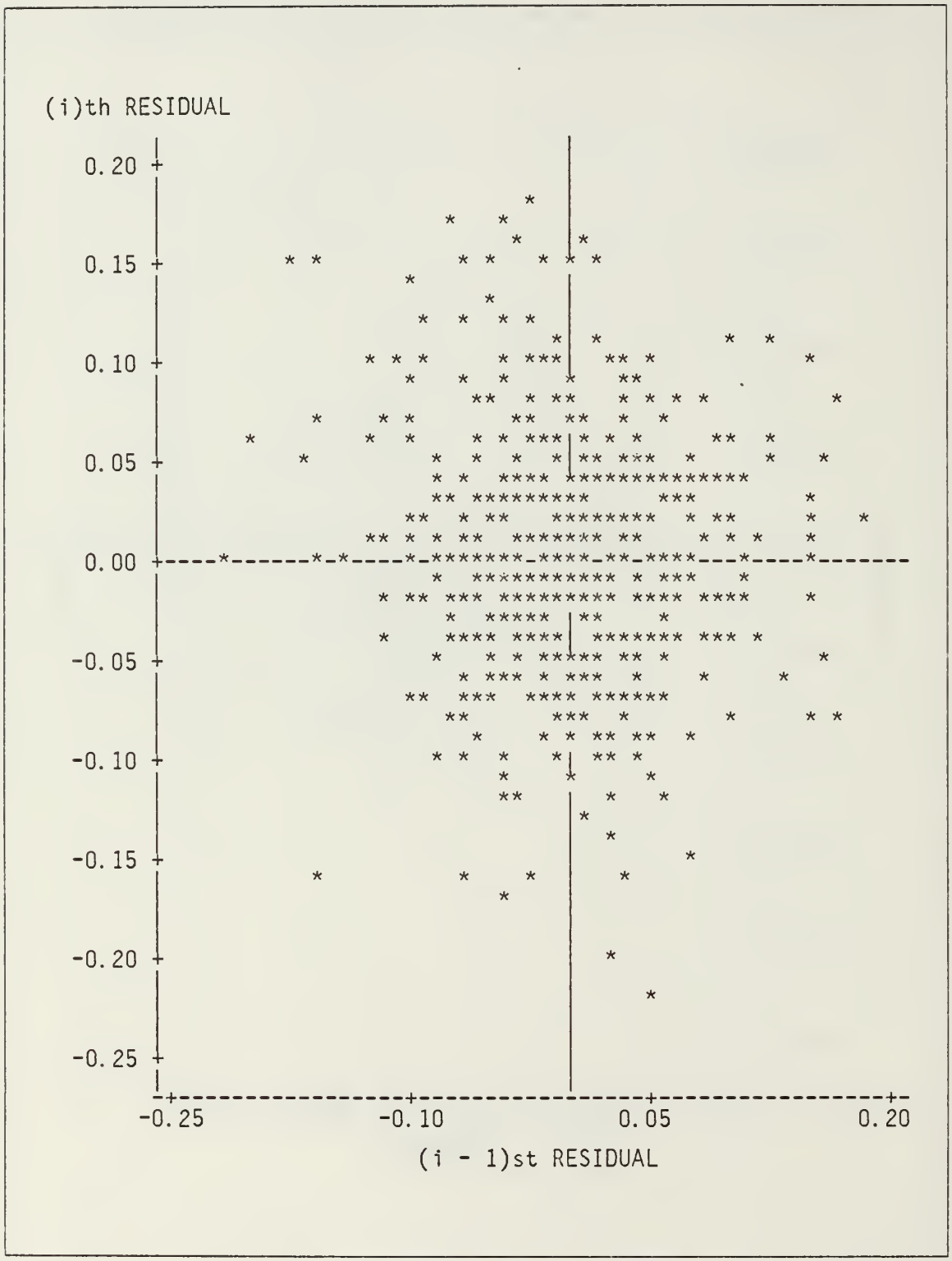


Figure 4.5 Residual Lag-1 Serial Correlation

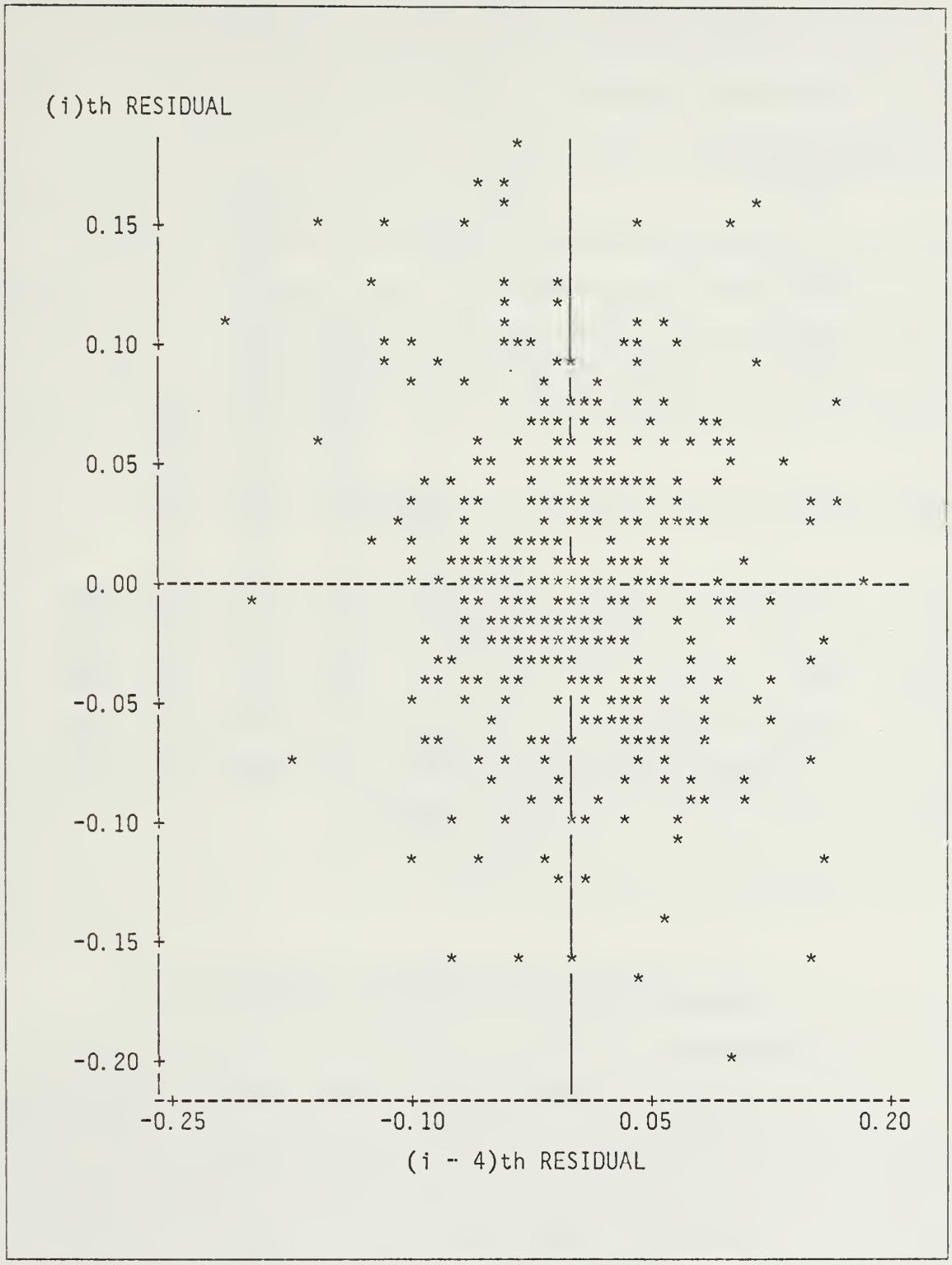


Figure 4.6 Residual Lag-4 Serial Correlation

respectively), but their estimated coefficients were consistently signed (always positive) and were frequently significant at levels just above the .15 acceptance threshold.

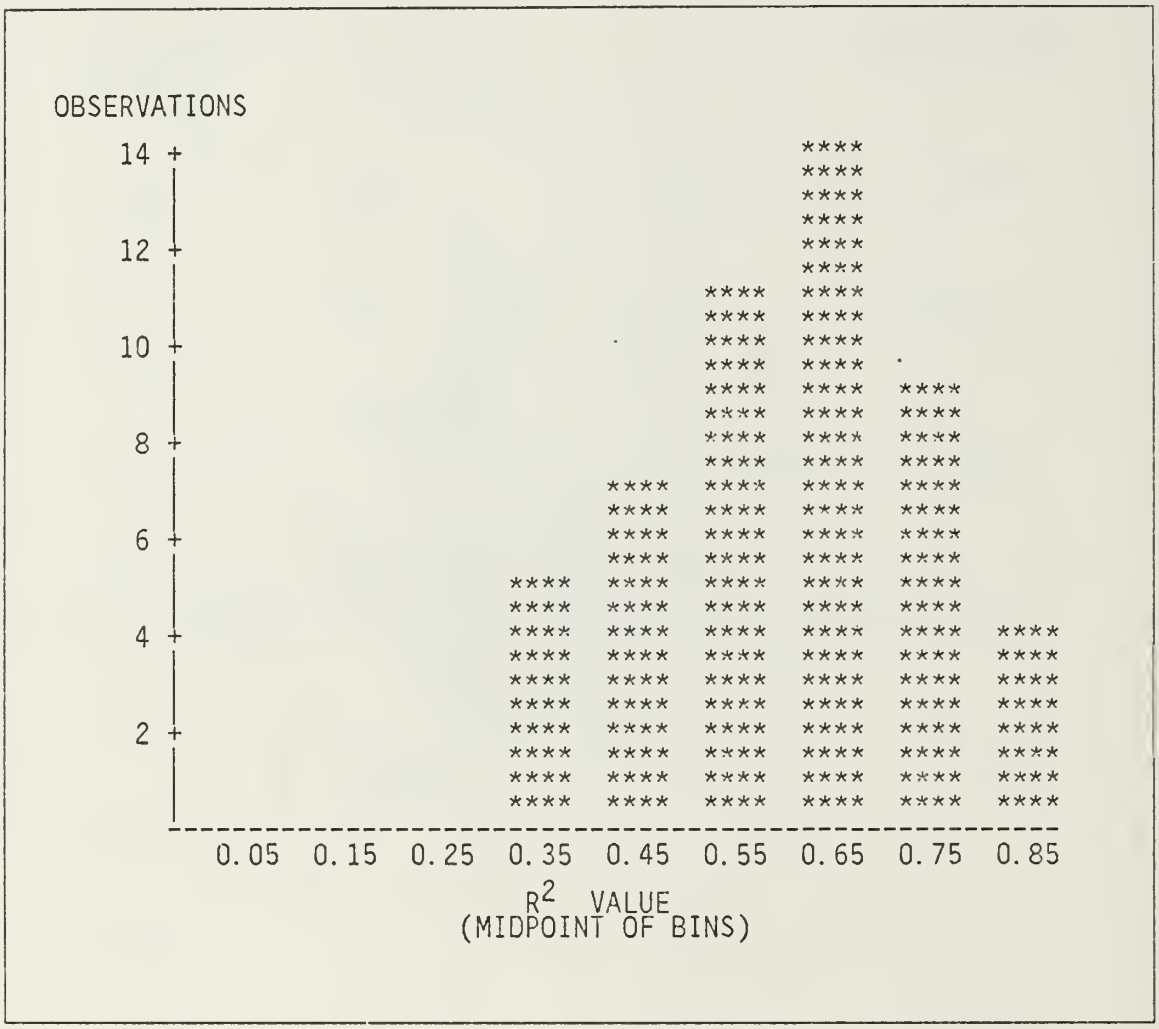


Figure 4.7 Distribution of R^2 Values
(Regression Procedure - Moderate Density MOS)

2. *The R^2 Statistic*

At Figure 4.7, the distribution of R^2 values, obtained from fitting the proposed overall model to the data available for the 50 moderate density MOS, is plotted, as previously, with a bar chart. Two points are worth noting with respect to Figure 4.7. First, as measured in terms of the R^2 statistic, our proposed overall model continues to serve us well in explaining the variation in retention rate through time at the MOS level. Second, the distribution of observations on the R^2 statistic for

moderate density MOS seems to be more highly spread than the R^2 distribution for high density MOS. This phenomenon is not unexpected when the smaller sample sizes associated with the moderate density MOS are considered. If our proposed overall model is correct, the decreased level of precision with which we can measure outcomes on the response variable, Y, will cause a general increase in the variability of the R^2 statistic, and a general decrease in it's mean value.

Our error term ε in the overall model actually accounts for the simultaneous effect of errors from several sources. The first, and most obvious source, is our inability to know or measure all factors which are critical to the reenlistment decision for all soldiers. A second significant source is our inability to measure the true response variable, Y. Recall, we estimate the zone A retention rate of a particular MOS for a particular quarter with:

$\hat{Y}$ = number of soldiers reenlisting for their own MOS / number of soldiers eligible to do so.

We have shown that the variance of the estimate generally increases with decreasing sample size. However for a particular MOS, if the general size of the sample can be considered stable in our period of study, then this measurement error is simply absorbed in the error term ε, without effect on the modelling assumptions. To the extent that the R^2 statistic can be thought of as the ratio of the variation in the data around $\overline{Y}$ explained by the regression, to total variation in the data around $\overline{Y}$ (which includes the variation accounted for by the error term), the decrease in the mean R^2 outcome, and increase in variability, are expected for the lower density MOS. [Ref. 8: pp. 93-94].

3. *Residual Analysis*

An extensive analysis of aggregate residual plots is not presented here because the results are very similar to the results we obtained when the overall model was fitted to the data for high density MOS. One plot which is worthy of note however, is the plot of residuals vs. sequence of observation at Figure 4.8. In our earlier analysis of residuals for high density MOS, we noted that residuals for quarters 6 and 9 appeared to be skewed positive. We note that for residuals associated with fitting the overall model to data for the 50 moderate density MOS, this perceived skewing is not apparent. This observation lessens our concern that our error term contains systematic and biasing components.

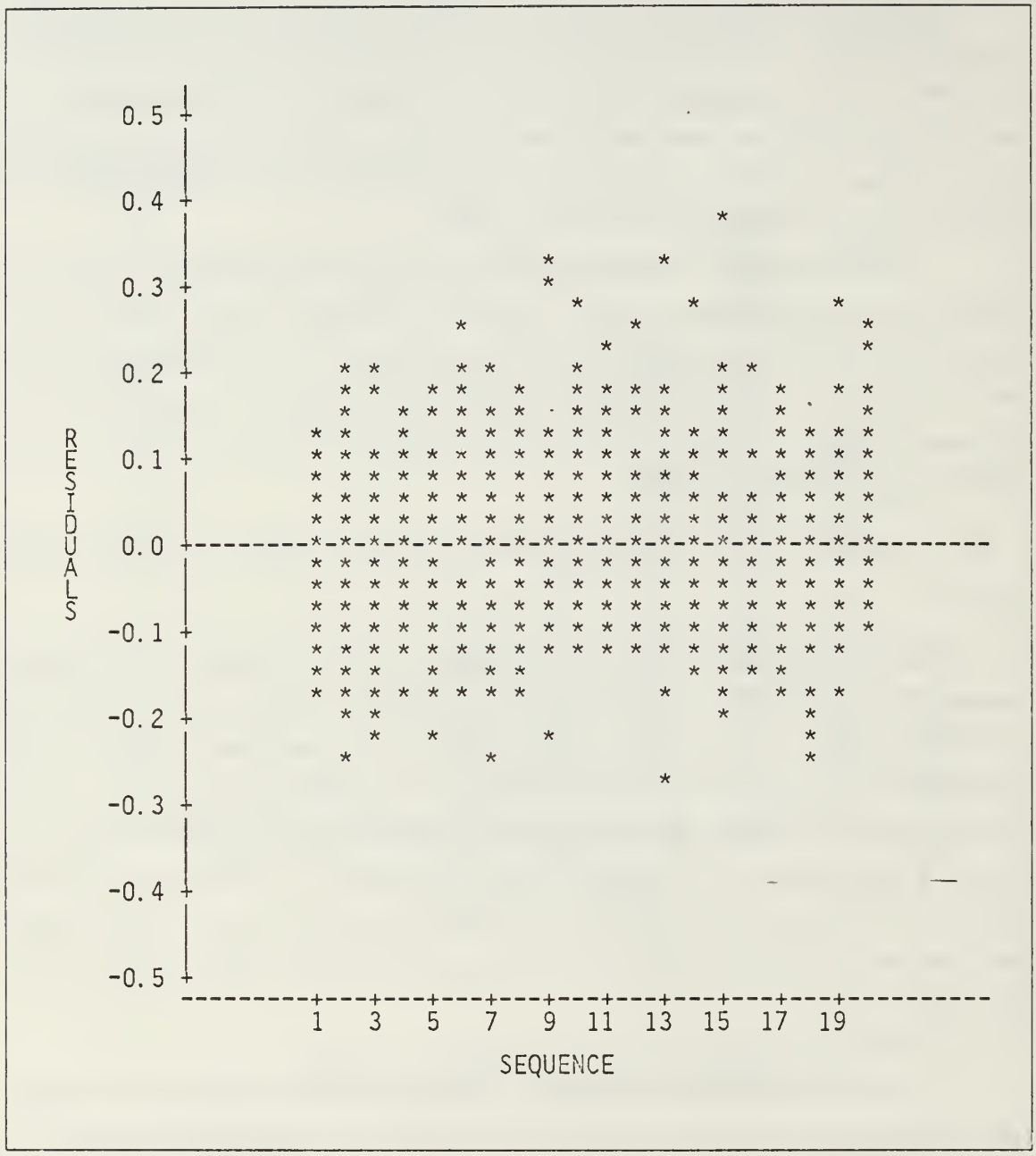


Figure 4.8 Residuals vs. Time Sequence
(Moderate Density MOS)

D. DATA TRANSFORMATIONS

It is standard practice in regression analysis to consider variance stabilizing transformations, such as the arcsin transformation, when the response variable is a parameter estimated by proportional data [Ref. 4: pp. 236-240]. Such a transformation

is considered because proportional type data typically do not have a uniform variance; the estimated variance of the data is dependent on the rate itself. However, these transformations are not used when the value of the estimated rates are in the range (0.3 - 0.7). In this range, the most common variance stabilizing transformations are nearly linear, and the dependence of the sample variance on the estimated rate is minimal. In the graph at Figure 4.3, we see no evidence which warrants a variance stabilizing transformation of our data. The overall model without transformation is believed best suited to the needs of our intended user.

E. A DEMONSTRATION OF MODEL USE

We have shown in Chapter III that, given our model is correct, a prediction of Y at X_0 is given by:

$$\hat{Y}_0 = b'X_0 \tag{4.2}$$

with variance given by:

$$V(\hat{Y}_0) = X_0' (X'X)^{-1} X_0 \sigma^2. \tag{4.3}$$

Using the error mean square term as our best estimate of σ^2, we can construct a 90% confidence interval for the true mean value of Y at X_0 as follows:

$$\hat{Y}_0 \ (+/-) \ 1.771(s)(X_0' (X'X)^{-1} X_0)^{1/2} \tag{4.4}$$

where s represents the square root of error mean square.

To demonstrate the use of this model, we have arbitrarily selected MOS 11H for the purpose of conducting sensitivity analysis. The value of the R^2 statistic when the overall model was fitted is .7283, and the variables QTRSQ, SRB, and REAL are significant at the .15 level of acceptance. Analysis of the residuals reveals no significant departure from normality.

To perform our analysis, we again resort to the SAS statistical software package. The $(X'X)^{-1}$ matrix, the estimated regression coefficients and the error mean square, calculated using PROC REG, were printed to an output file. The computational formulas shown in equations 4.2, 4.3, and 4.4 were added to this file, and it was

prepared as an input file to the SAS PROC MATRIX routine. Copies of the input files and output files involved in this procedure are at Appendix F.

The 7 dimensional vector of values on the independent variables at which we wish to predict an outcome for the dependent variable Y is represented by X_0. Let us hypothesize an X_0 value of (1, 2, 4, 0, 0.4, 0.25, 3.0), where the first position of the vector is reserved for the unity multiplier of the intercept term and the remaining values represent outcomes on the independent variables QTR, QTRSQ, SRB, RACE, DEP, and REAL respectively. The 90% confidence interval on the true mean value of Y at X_0 are shown on the first line at Table 4. The 90% confidence intervals on the true mean value of Y at X_0 when the hypothesized value of SRB is changed to levels 1, 2, and 3 are shown on lines 2, 3, and 4 of Table 4 respectively.

TABLE 5
SENSITIVITY ANALYSIS FOR MOS 11H

SRB Level	.90 LB	$\hat{Y}_0$	.90 UB	s.e.(Predict)
0	.301	.426	.551	.0707
1	.376	.475	.574	.0561
2	.433	.524	.615	.0511
3	.469	.573	.677	.0585

In the results summarized at Table 4, we observe two phenomena. First, and as expected, the value of $\hat{Y}_0$ increases at a steady rate of .049 with each unit increase in SRB level (.049 is the value of b_1). Second, and more importantly however, note the behavior of the standard error of the prediction (s.e. (Predict) - the square root of our $\hat{V}(\hat{Y}_0)$ term). It decreases through SRB level 2 and increases thereafter. This behavior is the result of our moving closer to the center of the sample data space. As we move further from the center of the sample data space, reliance on a point estimate for the response variable is increasingly dangerous. If we attempt to extrapolate beyond our sample data space, we can have very little confidence in the validity of our point prediction [Ref. 4: p. 8].

We note also that the widths of the 90% confidence intervals defined above can be approximately represented as $\hat{Y}_0$ (+/-) 10%. When MOS for which the overall model provided a better fit were considered (such as MOS 63B), these confidence intervals were more nearly approximated by $\hat{Y}_0$ (+/-) 3%.

It is not a simple matter to measure 6 dimensional data spaces. For our uses however, it is a simple enough matter to ensure that any sensitivity analysis conducted with respect to any particular independent variable, or combination of independent variables, remains in the range of values defined by the sample data space for those variables. In general, when the the independent variables are unrelated, the bulk of the potential problems associated with prediction are avoided if the sensitivity analysis is conducted within the individual value ranges of the independent variables.

We must be particularly careful when the estimated coefficient of any carrier variable, such as SRB, is interpreted as the effect of varying the level of the associated variable while the other values are unchanged. Even when that variable is unrelated to it's costock, the range of values for which such an interpretation is valid, as described by the sample data space, should be respected. This is best shown by example.

At Appendix E, we have examined the model parameter estimates for MOS 63B. We note that the coefficient for carrier variable SRB is estimated as .304. This estimate is based on a sample data range of (0, 1) for the variable SRB. Clearly, it is not reasonable to use this estimate as an effectiveness coefficient at SRB levels 2, 3 or higher (implying a 60%, 90%, or higher increase in retention rate over the SRB level 0 rate). Alternatives for prediction and comparison when we wish to extrapolate beyond our data space are described in the next section.

F. ALTERNATE MODELLING STRATEGIES

In developing the overall model, we considered a data base representing 24 high density MOS, which were authorized as of 30 September 1985, and for which an active SRB history existed during our period of analysis. We then fit the proposed overall model to 50 moderate density MOS with active SRB histories to verify our modelling assumptions. Based on our preceding analysis, we propose that the overall model be extended for general use in explaining the variation in zone A retention behavior for all MOS. We acknowledge however, that as the density associated with an MOS decreases, so does our ability to maintain small confidence intervals about our parameter and prediction estimates. As stated earlier, this is a consequence of

including the additional imprecision associated with our measurement of Y in the error term ε. If, in our examination of residuals however, we find no reason to discount our modelling assumptions, and no intuitive reason exists to discount these assumptions, then there is no reason to believe a better model exists.

In the event that the model suffers grossly from lack of fit, or other factors exist which cast doubt on the applicability of the model to a particular MOS, use of this model in a predictive procedure for that MOS is not advised. This situation is most likely to occur in fitting the model to data associated with very low density, highly technical MOS. In such a case, it is advisable to construct and maintain an MOS specific predictive model. Any inter-MOS comparison of the estimated coefficients of like carrier variables should not include these unique specialties.

Suggestions for using the developed overall model under extraordinary circumstances follow.

1. *Modelling a new MOS*

Typically, when a new MOS is introduced, personnel are reclassified from some other specialty, which is in turn reduced in size or eliminated. A pseudo-historic data base for the new MOS can then be created by including the records of the individual reenlistment decisions and SRB histories applicable to soldiers in the losing MOS.

2. *Modelling a Low Density MOS*

When the sample sizes involved in a very low density MOS are so small that acceptablely reliable estimates of the regression parameters cannot be attained, but the model is believed adequate, then it is recommended that the estimated coefficients of a like MOS, for which an adequate sample size is available, be used in retention rate prediction. This alternative is suggested in preference to grouping these low density MOS for two reasons. First, an explicit decision is made by the SRB program manager, as to which MOS can best represent the MOS of concern in retention rate projection. With the group method, we average the effects of several MOS. It is intuitive that our results with a single most similar MOS should be better. Second, we need not develop imaginative ways to group MOS unique factors, such as SRB level, across many MOS.

3. *Extrapolating Beyond the Sample Data Space.*

If the extrapolation is not too distant from the sample data space and does not involve extrapolating the SRB level, then it is recommended that we use the

developed model without modification, making clear our concern over the increasing danger of using a point prediction. If the extrapolation does involve the SRB variable, or the extrapolation is far beyond the data space described by our available data in any dimension, then selection of a like MOS, with a data space accommodating our needs, is recommended for use in analysis.

V. CONCLUSIONS AND RECOMMENDATIONS

In this thesis, the problem of developing a predictive model which explains the variation in zone A enlisted retention rates at the MOS level is formulated and solved using stepwise and ordinary least squares linear regression analysis. Inasmuch as the principle use of this model will be in the management of the SRB program, SRB level was initially included as a candidate carrier variable. Two other categories of candidate carrier variables were also included. The *endogenous* variables represent a demographic profile of an eligible reenlistment population. The *exogenous* variables represent the alternate career opportunities as perceived by the reenlistment decision-maker. This approach represents a significant improvement over earlier efforts to solve this problem, in that a capability to include a demographic profile of the eligible populations was not previously available to the analyst.

To allow for the inter-MOS comparison of the estimated regression coefficients associated with the SRB variable, an overall projection model, applicable to all MOS, was developed. We selected 24 high density MOS, which had active SRB histories in our sample period, to include in our initial analysis. Stepwise multiple linear regression analysis was used to find a best overall explanatory model, which could be used to project retention at the MOS level. The proposed overall model follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5 + \varepsilon \qquad (5.1)$$

where:

Y = retention rate

X_1 = SRB level

X_2 = fiscal quarter

X_3 = rate representing the race profile of an eligible population

X_4 = rate representing the dependent profile of an eligible population

X_5 = rate representing the real change in a soldier's pay through time

ε = error component with assumed distribution $N(0, \sigma^2)$

and β is a vector of the parameters to be estimated.

We note that the X_3 variable is included to account for the effects of the observed seasonal behavior in the retention rate. We note also that no variable is included in the proposed overall model which accounts for the effects of promotion program management.

Personnel inventory managers at MILPERCEN view the Army promotion program as a force alignment tool in the same way that accession and reclassification programs are viewed. Promotion opportunity to grades E5 and E6 are managed at the MOS level with the intention of providing incentives (or disincentives) for zone A soldiers to reenlist for their entry MOS. In not including an independent variable in our proposed overall model to account for this mechanism, we make no conclusions as to it's effectiveness, but we do conclude that the statistic provided us to measure it's effect is inadequate for that purpose. The measure preferred by the MILPERCEN program managers, promotion cut-off score, was unavailable during the period of our analysis. We recommend that an analysis similar to the one described in this report, including the promotion cut-off scores, be performed when a sufficient base of historic records are available.

We selected 50 moderate density MOS, which had active SRB histories in our sample period, to include in our validation analysis. Our results from this analysis were very favorable. We recommend the proposed model for use in predicting retention for all MOS with the following caveats:

1 Care must be taken in extrapolating beyond the region defined by our sample space.

2 Reliance on point estimates for retention become increasingly dangerous as the density of the MOS decreases.

3 When the estimated regression coefficient of the SRB variable is interpreted as the effect of varying the SRB level while the level of all other factors remains unchanged, the range of values for which that interpretation is valid must be respected.

When the regression coefficients cannot be reliably estimated from the available data, we recommend the use of the estimated regression coefficients of a like and more reliably modelled MOS. We prefer this alternative to the method of creating MOS groups, as has been done in past studies, for two reasons. First, a decision is explicitly made by the SRB program manager, as to which particular MOS model can best represent the MOS of concern. Second, the problems associated with grouping MOS unique data, such as SRB level, are avoided.

The actual estimated regression coefficients developed for each MOS in our analysis have not been included in this report. Instead, this analysis has been conducted using only those programming languages and analytic software available to the DCSPLANS, MILPERCEN, Force Plans Branch. All program code required to implement the analytic processes described in this report are included as appendices and referenced as appropriate.

It is recommended that the regression coefficients be estimated on a periodic basis using the programs and procedures described in Chapter IV of this report. If it becomes apparent that the overall model is no longer adequate, either through examination of the residuals or because some measureable factor not included in the overall model has become critical to the reenlistment decision (as could occur with a change to the EPMS), then we recommend that a zone A retention model be newly developed following the procedures set forth in Chapter III of this study.

APPENDIX A

FORTRAN PROGRAM TO PRODUCE DEMOGRAPHIC RATES

```
C************* VARIABLE DECLARATIONS ************************************
C
      INTEGER REUP,LEVEL,TERM,BASDY,BASDM,DEP,AFOT,TOTREC,TOTMOS,QTR,
     1A,B,C,TIS,Z,O,P,RECTOT(5,250,20),RACEY(5,250,20),DEPY(5,250,20),
     1SEXY(5,250,20),CIVEDY(5,250,20),AFOTY(5,250,20),REUPY(5,250,20),
     1TERMY(5,250,20),OTHER(5,20),REUPO(5,20),EDATEY,EDATEM
C
C
      REAL REUPR(5,250,20),RACER(5,250,20),DEPR(5,250,20),
     1SEXR(5,250,20),CIVEDR(5,250,20),AFQTR(5,250,20),TERMR(5,250,20)
C
C
      CHARACTER*1 PMOS*3,RACE,MARST,SEX,CIVED,TGTMOS(250)*3
C
C
      TOTREC = 0
C
C*************** READ MOS TARGETS ***************************************
C
      DO 5 I = 1,250
         READ(5,101,END=9) TGTMOS(I)
  5   CONTINUE
  101 FORMAT(A3)
C
C
  9   TOTMOS = I-1
C
C**************** INITIALIZATION ****************************************
C
      DO 10 A=1,5
         DO 11 B=1,250
            DO 12 C=1,20
               RECTOT(A,B,C)=0
               DEPY(A,B,C)=0
               SEXY(A,B,C)=0
               CIVEDY(A,B,C)=0
               AFOTY(A,B,C)=0
               REUPY(A,B,C)=0
               TERMY(A,B,C)=0
               OTHER(A,C)=0
               REUPO(A,C)=0
C
C
               RACER(A,B,C)=0.0
               DEPR(A,B,C)=0.0
               SEXR(A,B,C)=0.0
               CIVEDR(A,B,C)=0.0
               AFQTR(A,B,C)=0.0
               REUPR(A,B,C)=0.0
               TERMR(A,B,C)=0.0
  12  CONTINUE
  11  CONTINUE
  10  CONTINUE
C
C******** READ EACH RECORD (APPROX 481K) *********************************
C
  15  READ(11,102,END=19) PMOS,REUP,LEVEL,TERM,BASDY,BASDM,EDATEY,
     1EDATEM,RACE,MARST,DEP,SEX,CIVED,AFQT
  102 FORMAT(A3,3I1,4I2,2A1,I1,2A1,I2)
```

```
C
C
      TOTREC = TOTREC + 1
C
CCCC***************** ESTABLISH TIS ***************************************
C
      TIS = (EDATEY*12 + EDATEM) - (BASDY*12 + BASDM)
C
      IF(TIS.LT.21) THEN
         Z=1
      ELSE IF(TIS.LT.72) THEN
         Z=2
      ELSE IF(TIS.LT.120) THEN
         Z=3
      ELSE IF(TIS.LT.168) THEN
         Z=4
      ELSE
         Z=5
      ENDIF
C
CC**************** ESTABLISH QUARTER ************************************
C
      QTR = (((EDATEY*12 + EDATEM) - 970) / 3) + 1
C
C
      IF(QTR.LT.1 .OR. QTR.GT.20) GO TO 15
C
CCCC****************** START COUNT ***************************************
C
      DO 20 J=1,TOTMOS
         IF(PMOS.NE.TGTMOS(J)) THEN
            GO TO 20
         ELSE
            RECTOT(Z,J,QTR) = RECTOT(Z,J,QTR) + 1
            TERMY(Z,J,QTR) = TERMY(Z,J,QTR) + TERM
         ENDIF
C
C
         IF(RACE.NE.'C') THEN
         RACEY(Z,J,QTR) = RACEY(Z,J,QTR) + 1
C OTHER CODES: C,M(YELLOW),N,R(AMER IND),X,Z(UNK).
         ENDIF
C
         IF(DEP.GE.2) THEN
         DEPY(Z,J,QTR) = DEPY(Z,J,QTR) + 1
         ENDIF
C
         IF(SEX.EQ.'F') THEN
         SEXY(Z,J,QTR) = SEXY(Z,J,QTR) + 1
         ENDIF
C
         IF(CIVED.GT.'D') THEN
         CIVEDY(Z,J,QTR) = CIVEDY(Z,J,QTR) + 1
C OTHER CODES:0,1,2,3,4,5,6,7,8,A,B,C,D,//E...W,Y(NO Z).
         ENDIF
C
         IF(AFQT.LT.50) THEN
         AFQTY(Z,J,QTR) = AFQTY(Z,J,QTR) + 1
C BRKPTS: 4(16-30),3B(31-49),3A(50-64),2(65-92),1(93-99)
         ENDIF
C
         IF(REUP.EQ.1) THEN
         REUPY(Z,J,QTR) = REUPY(Z,J,QTR) + 1
         ENDIF
C
```

```
         GO TO 15
   20  CONTINUE
C
C
       IF(QTR.LT.1 .OR. QTR.GT.20) GO TO 15
          OTHER(Z,QTR) = OTHER(Z,QTR) + 1
       IF(REUP.EQ.1) THEN
          REUPO(Z,QTR) = REUPO(Z,QTR) + 1
       ENDIF
C
C
       GO TO 15
C
C************* DIVIDE TO GET RATES ************************************
C
   19  DO 30 L = 1,5
          DO 40 M = 1,TOTMOS
            DO 50 N = 1,20
            IF(RECTOT(L,M,N).LT.1) GO TO 50
            RACER(L,M,N) = (FLOAT(RACEY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            DEPR(L,M,N) = (FLOAT(DEPY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            SEXR(L,M,N) = (FLOAT(SEXY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            CIVEDR(L,M,N) = (FLOAT(CIVEDY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            AFQTR(L,M,N) = (FLOAT(AFQTY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            REUPR(L,M,N) = (FLOAT(REUPY(L,M,N)))/(FLOAT(RECTOT(L,M,N)))
            IF(REUPY(L,M,N).LT.1) REUPY(L,M,N) = 100000
            TERMR(L,M,N) = (FLOAT(TERMY(L,M,N)))/(FLOAT(REUPY(L,M,N)))
            IF(REUPY(L,M,N).EQ.100000) REUPY(L,M,N) = 0
   50  CONTINUE
   40  CONTINUE
   30  CONTINUE
C
C***************** OUTPUT *************************************************
C
C
       DO 60 O=1,TOTMOS
          WRITE(13,514) (REUPY(2,O,P),P=1,20)
   60  CONTINUE
       DO 61 O=1,TOTMOS
          WRITE(13,513) (REUPR(3,O,P),P=1,20)
   61  CONTINUE
C
C
       DO 62 O=1,TOTMOS
          WRITE(13,514) (RACEY(2,O,P),P=1,20)
   62  CONTINUE
       DO 63 O=1,TOTMOS
          WRITE(13,513) (RACER(3,O,P),P=1,20)
   63  CONTINUE
C
C
       DO 64 O=1,TOTMOS
          WRITE(13,514) (DEPY(2,O,P),P=1,20)
   64  CONTINUE
       DO 65 O=1,TOTMOS
          WRITE(13,513) (DEPR(3,O,P),P=1,20)
   65  CONTINUE
C
C
       DO 66 O=1,TOTMOS
          WRITE(13,514) (SEXY(2,O,P),P=1,20)
   66  CONTINUE
       DO 67 O=1,TOTMOS
          WRITE(13,513) (SEXR(3,O,P),P=1,20)
   67  CONTINUE
C
C
```

```
      DO 68 O=1,TOTMOS
         WRITE(13,514) (CIVEDY(2,O,P),P=1,20)
   68 CONTINUE
      DO 69 O=1,TOTMOS
         WRITE(13,513) (CIVEDR(3,O,P),P=1,20)
   69 CONTINUE
C
C
      DO 70 O=1,TOTMOS
         WRITE(13,514) (AFQTY(2,O,P),P=1,20)
   70 CONTINUE
      DO 71 O=1,TOTMOS
         WRITE(13,513) (AFQTR(3,O,P),P=1,20)
   71 CONTINUE
C
C
      DO 72 O=1,TOTMOS
         WRITE(13,514) (TERMY(2,O,P),P=1,20)
   72 CONTINUE
      DO 73 O=1,TOTMOS
         WRITE(13,513) (TERMR(3,O,P),P=1,20)
   73 CONTINUE
C
C
      DO 74 O=1,TOTMOS
         WRITE(13,514) (RECTOT(2,O,P),P=1,20)
   74 CONTINUE
      DO 75 O=1,TOTMOS
         WRITE(13,514) (RECTOT(3,O,P),P=1,20)
   75 CONTINUE
C
C
      DO 76 O=1,5
         WRITE(13,514) (OTHER(O,P),P=1,20)
   76 CONTINUE
C
C
      DO 77 O=1,5
         WRITE(13,514) (REUPO(O,P),P=1,20)
   77 CONTINUE
C
C
C****************** FORMATS ******************************************
C
  513 FORMAT(20(F5.3,1X))
  514 FORMAT(20(I5,1X))
      STOP
      END
```

APPENDIX B

CORRELATION MATRIX

	REUP	SRB	RACE	DEP	SEX	EDUCATE	AFQT
REUP	1.00000	0.40106	0.34487	0.24321	0.23577	-0.16155	0.19732
SRB	0.40106	1.00000	-0.16719	0.07366	-0.19467	-0.00109	-0.11239
RACE	0.34487	-0.16719	1.00000	-0.10544	0.52904	0.00509	0.55851
DEP	0.24321	0.07366	-0.10544	1.00000	-0.19403	0.01708	0.01983
SEX	0.23577	-0.19467	0.52904	-0.19403	1.00000	0.25536	0.00782
EDUCATE	-0.16155	-0.00109	0.00509	0.01708	0.25536	1.00000	-0.33314
AFQT	0.19732	-0.11239	0.55851	0.01983	0.00782	-0.33314	1.00000
E5TEST2	0.01142	0.35958	-0.21260	-0.06718	-0.27534	-0.12471	-0.16466
E6TEST2	-0.08781	0.30615	-0.29116	0.12697	-0.51307	-0.11160	-0.13838
QTR	-0.32714	-0.06137	-0.01095	-0.29468	-0.02192	0.21847	0.00070
UNEMPLY	0.17660	-0.21680	0.12927	-0.09619	0.07544	-0.61271	0.17733
REAL	0.25425	-0.10198	0.14387	-0.10925	0.05044	-0.24003	0.06349

	E5TEST2	E6TEST2	QTR	UNEMPLY	REAL
REUP	0.01142	-0.08781	-0.32714	0.17660	0.25425
SRB	0.35958	0.30615	-0.06137	-0.21680	-0.10198
RACE	-0.21260	-0.29116	-0.01095	0.12927	0.14387
DEP	-0.06718	0.12697	-0.29468	-0.09619	-0.10925
SEX	-0.27534	-0.51307	-0.02192	0.07544	0.05044
EDUCATE	-0.12471	-0.11160	0.21847	-0.61271	-0.24003
AFQT	-0.16466	-0.13838	0.00070	0.17733	0.06349
E5TEST2	1.00000	0.12996	-0.00004	-0.00011	-0.00010
E6TEST2	0.12996	1.00000	0.00005	-0.00007	-0.00008
QTR	-0.00004	0.00005	1.00000	-0.02815	0.00000
UNEMPLY	-0.00011	-0.00007	-0.02815	1.00000	0.62229
REAL	-0.00010	-0.00008	0.00000	0.62229	1.00000

APPENDIX C

SAMPLE INPUT FILE - SAS PROC STEPWISE

----------MOS=63B----------

OBS	REUP	SRB	RACE	DEP
1	0.565200	1	0.315900	0.204300
2	0.464800	0	0.344600	0.207600
3	0.356500	0	0.325200	0.161900
4	0.286700	0	0.359400	0.165000
5	0.382400	0	0.398200	0.199100
6	0.680100	0	0.454000	0.211400
7	0.488900	0	0.388200	0.233400
8	0.403300	0	0.418300	0.201600
9	0.532600	0	0.436400	0.245700
10	0.554500	0	0.397200	0.260700
11	0.416500	0	0.399600	0.210400
12	0.293900	0	0.385900	0.175800
13	0.422500	0	0.399200	0.246100
14	0.523500	0	0.395500	0.222200
15	0.394400	0	0.380500	0.228600
16	0.270500	0	0.384100	0.209100
17	0.345300	0	0.351300	0.225500
18	0.312400	0	0.325300	0.249500
19	0.367100	0	0.384900	0.170600
20	0.247700	0	0.361700	0.165700

OBS	SEX	EDUCATE	AFQT	E5TEST2
1	0.0377000	0.782600	0.636200	0.1790
2	0.0392000	0.797700	0.674900	0.2210
3	0.0435000	0.789100	0.650300	0.1120
4	0.0350000	0.888100	0.664300	-0.0710
5	0.0407000	0.834800	0.653800	-0.1250
6	0.0368000	0.829000	0.704000	0.2080
7	0.0491000	0.909100	0.653600	-0.3500
8	0.0354000	0.862400	0.666200	-0.2920
9	0.0430000	0.773200	0.721600	-0.2170
10	0.0248000	0.773800	0.732400	-0.3000
11	0.0381000	0.816100	0.683900	-0.4960
12	0.0505000	0.851500	0.643400	-0.5790
13	0.0988000	0.829500	0.676400	-0.6040
14	0.0885000	0.811700	0.655400	-0.5040
15	0.0628000	0.844700	0.713800	-0.3960
16	0.0420000	0.902300	0.622700	-0.2830
17	0.0659000	0.924200	0.592800	-0.3830
18	0.0813000	0.883500	0.659900	-2.6880
19	0.0496000	0.932500	0.730200	-1.3460
20	0.0395000	0.952900	0.597300	-1.2960

OBS	E6TEST2	QTR	UNEMPLY	REAL
1	-1.1250	1	7.4000	0.80000
2	-1.5960	2	7.4000	0.80000
3	-1.5290	3	7.4000	0.80000
4	-1.2330	4	8.2000	0.80000
5	-1.0130	1	8.8000	9.30000
6	-0.9500	2	9.5000	9.30000
7	-1.0830	3	9.9000	9.30000
8	-1.2880	4	10.6000	9.30000
9	-1.4250	1	10.4000	1.10000
10	-1.4250	2	10.1000	1.10000
11	-1.3670	3	9.3000	1.10000
12	-0.6080	4	8.5000	1.10000
13	-0.7540	1	7.9000	-0.20000
14	-0.6380	2	7.5000	-0.20000
15	-0.6290	3	7.4000	-0.20000
16	-0.7420	4	7.2000	-0.20000
17	0.1500	1	7.3000	0.80000
18	-0.4580	2	7.3000	0.80000
19	-0.3790	3	7.2000	0.80000
20	-0.3830	4	7.0000	0.80000

APPENDIX D

SAMPLE OUTPUT FILE - SAS PROC STEPWISE

MOS=63B

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE REUP

STEP 1 VARIABLE EDUCATE ENTERED R SQUARE = 0.40205418
C(P) = 45.63109799

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	0.09995765	0.09995765	12.10	0.0027
ERROR	18	0.14865970	0.00825887		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1.52788549				
EDUCATE	-1.30962992	0.37644450	0.09995765	12.10	0.0027

BOUNDS ON CONDITION NUMBER: 1, 2

STEP 2 VARIABLE REAL ENTERED R SQUARE = 0.54491298
C(P) = 32.90644519

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	0.13547482	0.06773741	10.18	0.0012
ERROR	17	0.11314253	0.00665544		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1.54759367				
EDUCATE	-1.36639281	0.33882394	0.10823804	16.26	0.0009
REAL	0.01207975	0.00522910	0.03551718	5.34	0.0337

BOUNDS ON CONDITION NUMBER: 1.005287, 8.042296

STEP 3 VARIABLE QTRSQ ENTERED R SQUARE = 0.64745038
C(P) = 24.33777424

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	0.16096740	0.05365580	9.79	0.0007
ERROR	16	0.08764995	0.00547812		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1.27877948				
EDUCATE	-0.98469643	0.35468525	0.04222311	7.71	0.0135
QTRSQ	-0.00725385	0.00336262	0.02549258	4.65	0.0465
REAL	0.01165255	0.00474824	0.03299198	6.02	0.0260

BOUNDS ON CONDITION NUMBER: 1.338361, 22.06034

STEP 4 VARIABLE RACE ENTERED R SQUARE = 0.69900259
C(P) = 21.02421694

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	0.17378417	0.04344604	8.71	0.0008
ERROR	15	0.07483318	0.00498888		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	0.88054612				
RACE	0.83180763	0.51896133	0.01281677	2.57	0.1298
EDUCATE	-0.87244692	0.34564568	0.03178474	6.37	0.0234
QTRSQ	-0.00776269	0.00322462	0.02891151	5.80	0.0294
REAL	0.00758029	0.00519492	0.01062225	2.13	0.1651

BOUNDS ON CONDITION NUMBER: 1.395655, 43.2307

STEP 5 VARIABLE REAL REMOVED R SQUARE = 0.65627728
C(P) = 23.42797351

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	0.16316192	0.05438731	10.18	0.0005
ERROR	16	0.08545543	0.00534096		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	0.68753424				
RACE	1.20215572	0.46836293	0.03518650	6.59	0.0207
EDUCATE	-0.78645354	0.35239792	0.02660107	4.98	0.0403
QTRSQ	-0.00815958	0.00332457	0.03217236	6.02	0.0259

BOUNDS ON CONDITION NUMBER: 1.355083, 22.25706

STEP 6 VARIABLE SRB ENTERED R SQUARE = 0.74588024
C(P) = 16.19247340

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	0.18543877	0.04635969	11.01	0.0002
ERROR	15	0.06317858	0.00421191		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	0.30239640				
SRB	0.18328386	0.07969602	0.02227685	5.29	0.0362
RACE	1.71637743	0.47221401	0.05564504	13.21	0.0024
EDUCATE	-0.58192906	0.32533230	0.01347610	3.20	0.0939
QTRSQ	-0.00726620	0.00297778	0.02507889	5.95	0.0276

BOUNDS ON CONDITION NUMBER: 1.464518, 44.55447

STEP 7 VARIABLE QTR ENTERED R SQUARE = 0.88534414
C(P) = 3.81773702

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	0.22011191	0.04402238	21.62	0.0001
ERROR	14	0.02850544	0.00203610		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.27080045				
SRB	0.27969908	0.06013548	0.04404739	21.63	0.0004
RACE	2.04097646	0.33761270	0.07441116	36.55	0.0001
EDUCATE	-0.34220227	0.23353805	0.00437170	2.15	0.1649
QTR	0.23168421	0.05614352	0.03467315	17.03	0.0010
QTRSQ	-0.05231981	0.01111232	0.04513596	22.17	0.0003

BOUNDS ON CONDITION NUMBER: 39.11734, 824.5804

STEP 8 VARIABLE EDUCATE REMOVED R SQUARE = 0.86776010
C(P) = 3.63014835

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	0.21574022	0.05393505	24.61	0.0001
ERROR	15	0.03287713	0.00219181		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.62921701				
SRB	0.30971234	0.05866171	0.06109564	27.87	0.0001
RACE	2.18472453	0.33517003	0.09312487	42.49	0.0001
QTR	0.25214804	0.05641976	0.04377755	19.97	0.0005
QTRSQ	-0.05759786	0.01090687	0.06112444	27.89	0.0001

BOUNDS ON CONDITION NUMBER: 36.30779, 592.6314

STEP 9 VARIABLE E5TEST2 ENTERED R SQUARE = 0.91278288
C(P) = 0.98958872

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	0.22693366	0.04538673	29.30	0.0001
ERROR	14	0.02168369	0.00154883		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.51081673				
SRB	0.26583755	0.05194297	0.04056805	26.19	0.0002
RACE	1.90943944	0.29978309	0.06283515	40.57	0.0001
E5TEST2	0.03974190	0.01478323	0.01119344	7.23	0.0177
QTR	0.25877750	0.04749181	0.04598548	29.69	0.0001
QTRSQ	-0.05890404	0.00918143	0.06374917	41.16	0.0001

BOUNDS ON CONDITION NUMBER: 36.40594, 758.1488

STEP 10 VARIABLE UNEMPLY ENTERED R SQUARE = 0.92644837
C(P) = 1.58106751

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	6	0.23033114	0.03838852	27.29	0.0001
ERROR	13	0.01828621	0.00140663		
TOTAL	19	0.24861735			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.49008893				
SRB	0.25534625	0.04995923	0.03674582	26.12	0.0002
RACE	1.52242447	0.37898701	0.02269882	16.14	0.0015
E5TEST2	0.03637129	0.01425421	0.00915823	6.51	0.0241
QTR	0.25377900	0.04537328	0.04400392	31.28	0.0001
QTRSQ	-0.05800326	0.00876897	0.06154427	43.75	0.0001
UNEMPLY	0.01577333	0.01014928	0.00339748	2.42	0.1442

BOUNDS ON CONDITION NUMBER: 36.58979, 952.6237

NO OTHER VARIABLES MET THE 0.1500 SIGNIFICANCE LEVEL FOR ENTRY

SUMMARY OF STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE REUP

STEP	VARIABLE ENTERED	VARIABLE REMOVED	NUMBER IN	PARTIAL R**2	MODEL R**2	C(P)
1	EDUCATE		1	0.4021	0.4021	45.6311
2	REAL		2	0.1429	0.5449	32.9064
3	QTRSQ		3	0.1025	0.6475	24.3378
4	RACE		4	0.0516	0.6990	21.0242
5		REAL	3	0.0427	0.6563	23.4280
6	SRB		4	0.0896	0.7459	16.1925
7	QTR		5	0.1395	0.8853	3.8177
8		EDUCATE	4	0.0176	0.8678	-3.6301
9	E5TEST2		5	0.0450	0.9128	0.9896
10	UNEMPLY		6	0.0137	0.9264	1.5811

STEP	VARIABLE ENTERED	VARIABLE REMOVED	F	PROB>F
1	EDUCATE		12.1031	0.0027
2	REAL		5.3366	0.0337
3	QTRSQ		4.6535	0.0465
4	RACE		2.5691	0.1298
5		REAL	2.1292	0.1651
6	SRB		5.2890	0.0362
7	QTR		17.0292	0.0010
8		EDUCATE	2.1471	0.1649
9	E5TEST2		7.2270	0.0177
10	UNEMPLY		2.4153	0.1442

APPENDIX E

SAMPLE OUTPUT FILE - SAS PROC REG

DEP VARIABLE: REUP

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	6	0.21740661	0.03623444	15.092	0.0001
ERROR	13	0.03121074	0.002400826		
C TOTAL	19	0.24861735			

ROOT MSE	0.04899822	R-SQUARE	0.8745
DEP MEAN	0.41544	ADJ R-SQ	0.8165
C.V.	11.7943		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER=0	PROB > |T|
INTERCEP	1	-0.59285005	0.18763783	-3.160	0.0075
QTR	1	0.24775924	0.05948661	4.165	0.0011
QTRSQ	1	-0.05638160	0.01178244	-4.785	0.0004
SRB	1	0.30434495	0.06217371	4.895	0.0003
RACE	1	2.00096212	0.41845916	4.782	0.0004
DEP	1	0.13621363	0.50887958	0.268	0.7932
REAL	1	0.002993859	0.003637893	0.823	0.4254

OBS	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	LOWER95% PREDICT
1	0.5652	0.5652	0.0490	0.4593	0.6711	0.4155
2	0.4648	0.3973	0.0215	0.3509	0.4438	0.2817
3	0.3565	0.3182	0.0304	0.2526	0.3837	0.1936
4	0.2867	0.2401	0.0237	0.1889	0.2913	0.1225
5	0.3824	0.4503	0.0356	0.3734	0.5272	0.3194
6	0.6801	0.6422	0.0339	0.5689	0.7155	0.5135
7	0.4889	0.4794	0.0327	0.4087	0.5501	0.3521
8	0.4033	0.3884	0.0322	0.3187	0.4580	0.2617
9	0.5326	0.5085	0.0306	0.4424	0.5746	0.3837
10	0.5545	0.5107	0.0246	0.4575	0.5639	0.3922
11	0.4165	0.4745	0.0202	0.4309	0.5182	0.3600
12	0.2939	0.2955	0.0225	0.2469	0.3441	0.1790
13	0.4225	0.4302	0.0267	0.3726	0.4878	0.3097
14	0.5235	0.4982	0.0212	0.4524	0.5440	0.3829
15	0.3944	0.4349	0.0210	0.3896	0.4802	0.3197
16	0.2705	0.2925	0.0265	0.2353	0.3497	0.1722
17	0.3453	0.3346	0.0287	0.2726	0.3965	0.2119
18	0.3124	0.3644	0.0325	0.2943	0.4346	0.2375
19	0.3671	0.4388	0.0273	0.3798	0.4978	0.3176
20	0.2477	0.2448	0.0234	0.1941	0.2954	0.1274

OBS	UPPER95% PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2
1	0.7149	1.5E-16	0	.	
2	0.5130	0.0675	0.0440	1.5322	***
3	0.4427	0.0383	0.0385	0.9969	*
4	0.3577	0.0466	0.0429	1.0865	**
5	0.5811	-0.0679	0.0337	-2.0164	****
6	0.7710	0.0379	0.0354	1.0715	**
7	0.6067	.0094981	0.0365	0.2606	
8	0.5151	0.0149	0.0369	0.4041	
9	0.6333	0.0241	0.0383	0.6294	*
10	0.6292	0.0438	0.0424	1.0334	**
11	0.5890	-0.0580	0.0446	-1.3000	**
12	0.4120	-.001592	0.0435	-0.0366	
13	0.5507	-.007735	0.0411	-0.1882	
14	0.6135	0.0253	0.0442	0.5730	*
15	0.5500	-0.0405	0.0443	-0.9146	*
16	0.4129	-0.0220	0.0412	-0.5344	*
17	0.4572	0.0107	0.0397	0.2699	
18	0.4914	-0.0520	0.0367	-1.4178	**
19	0.5600	-0.0717	0.0407	-1.7622	***
20	0.3621	.0029051	0.0430	0.0675	

OBS	COOK'S D
1	.
2	0.080
3	0.088
4	0.051
5	0.650
6	0.151
7	0.008
8	0.018
9	0.036
10	0.052
11	0.049
12	0.000
13	0.002
14	0.011
15	0.027
16	0.017
17	0.005
18	0.225
19	0.200
20	0.000

SUM OF RESIDUALS	2.49800E-16
SUM OF SQUARED RESIDUALS	0.03121074

APPENDIX F

SAMPLE INPUT / OUTPUT FILES - SAS PROC MATRIX

********** INPUT **********

X'XINV	COL1	COL2	COL3	COL4
	COL5	COL6	COL7	
ROW1	5.9325	-1.91717	0.301021	0.114193
	-1.67589	-16.3181	-0.0252033	
ROW2	-1.91717	1.37198	-0.258016	-0.0292765
	0.0739555	2.35476	0.00429438	
ROW3	0.301021	-0.258016	0.0509633	0.00705693
	-0.0205588	-0.219774	-.000129725	
ROW4	0.114193	-0.0292765	0.00705693	0.10918
	-0.407743	-0.585478	0.00624465	
ROW5	-1.67589	0.0739555	-0.0205588	-0.407743
	7.891	0.288322	-0.115236	
ROW6	-16.3181	2.35476	-0.219774	-0.585478
	0.288322	68.4104	0.155382	
ROW7	-0.0252033	0.00429438	-.000129725	0.00624465
	-0.115236	0.155382	0.0061736	

XO	COL1
ROW1	1
ROW2	2
ROW3	4
ROW4	0
ROW5	0.4
ROW6	0.25
ROW7	3

b'	COL1	COL2	COL3	COL4
	COL5	COL6	COL7	
	-0.104345	0.134489	-0.0313409	0.0491367
	0.32371	0.888201	0.0116723	

EMS	COL1
ROW1	0.00610136

TCRIT	COL1
ROW1	1.771

```
**********  OUTPUT 1  -  SRB LEVEL = 0  **********

     YO                    COL1
             ROW1        0.425821

     VAR(YO)               COL1
             ROW1       0.0050044

     CI(LOW)               COL1
             ROW1        0.300537

     CI(HIGH)              COL1
             ROW1        0.551105

**********  OUTPUT 2  -  SRB LEVEL = 1  **********

     YO                    COL1
             ROW1        0.474958

     VAR(YO)               COL1
             ROW1      0.00314623

     CI(LOW)               COL1
             ROW1         0.37562

     CI(HIGH)              COL1
             ROW1        0.574295

**********  OUTPUT 3  -  SRB LEVEL = 2  **********

     YO                    COL1
             ROW1        0.524094

     VAR(YO)               COL1
             ROW1      0.00262035

     CI(LOW)               COL1
             ROW1        0.433438
```

```
CI(HIGH)                COL1
        ROW1         0.614751

**********  OUTPUT 4  -  SRB LEVEL = 3  **********

YO                      COL1
        ROW1         0.573231

VAR(YO)                 COL1
        ROW1       0.00342676

CI(LOW)                 COL1
        ROW1         0.469559

CI(HIGH)                COL1
        ROW1         0.676903
```

APPENDIX G

EXTRACT OF SAS V5 PROGRAMMING COMMANDS USED IN THIS STUDY

```
OPTIONS LINESIZE=64
        PAGESIZE=60;
DATA ARRAY1;
INPUT
      MOS $ SRB REUP RACE DEP SEX EDUCATE AFQT TERM E5TEST1 E6TEST1
      E5TEST2 E6TEST2 QTR UNEMPLY REAL SEQ YEAR;
CARDS;

******   (include data arrays)   ******

;
PROC PRINT
      DATA=ARRAY1 N UNIFORM;
      VAR REUP SRB RACE DEP SEX EDUCATE AFQT E5TEST2 E6TEST2
      QTR UNEMPLY REAL;
      BY MOS;
PROC CORR DATA=ARRAY1 NOSIMPLE;
      VAR REUP SRB RACE DEP SEX EDUCATE AFQT E5TEST2 E6TEST2
      QTR UNEMPLY REAL;
PROC PRINT;
PROC STEPWISE DATA=ARRAY1;
      MODEL REUP = SRB SRB*2 RACE DEP SEX EDUCATE AFQT E5TEST1 E6TEST1
      E5TEST2 E6TEST2 TERM QTR QTR*2 SEQ YEAR UNEMPLY REAL
      / SLE=.150 SLS=.150;
      BY MOS;
PROC REG
      DATA=ARRAY1;
      MODEL REUP = QTR QTRSQ SRB RACE DEP REAL / I P R CLM CLI INFLUENCE;
      BY MOS;
      OUTPUT OUT=OUT1 P=YHAT1 R=RESID1;
PROC CHART
      DATA=OUT1;
      HBAR RESID1/MIDPOINTS=-.25 TO .25 BY .010;
PROC PLOT
      DATA=OUT1;
      PLOT RESID1*YHAT1='*' RESID1*SEQ='*'/VREF=0;
DATA OUT11;
      SET OUT1;
      IF SEQ=1 THEN DELETE;
      R11=RESID1;
DATA OUT41;
      SET OUT1;
      IF SEQ<=4 THEN DELETE;
      R41=RESID1;
DATA OUT12;
      SET OUT1;
      IF SEQ=20 THEN DELETE;
      R12=RESID1;
DATA OUT42;
      SET OUT1;
      IF SEQ>=17 THEN DELETE;
      R42=RESID1;
DATA LAG1;
      MERGE OUT11 OUT12;
```

```
DATA LAG4;
     MERGE OUT41 OUT42;
PROC PLOT
     DATA=LAG1;
     PLOT R11*R12='*' / VREF=0 HREF=0;
PROC PLOT
     DATA=LAG4;
     PLOT R41*R42='*' / VREF=0 HREF=0;
```

LIST OF REFERENCES

1. Rand Corporation, *Reenlistment Bonuses and First Term Retention*, September 1977.

2. Concepts Analysis Agency, *Selective Reenlistment Bonus Study*, September 1982.

3. Mosteller, F. and Tukey, J. W., *Data Analysis and Regression* Addison-Wesley Publishing Company, 1977.

4. Draper, N. R. and Smith, H., *Applied Regression Analysis*, 2d ed., John Wiley and Sons, 1981.

5. Mood, A. M., Graybill, F. A., and Boes, D. C., *Introduction to the Theory of Statistics*, 3rd ed., McGraw Hill Book Company, 1974.

6. SAS Institute Inc., *SAS User's Guide*: *Basics*, Version 5 ed., 1985.

7. SAS Institute Inc., *SAS User's Guide*: *Statistics*, Version 5 ed., 1985.

8. Neter, S. and Wasserman, W., *Applied Linear Statistical Models*, Richard D. Irwin Inc., 1974.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginina 22304-6145	2
2.	Library,Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Deputy Underseeretary of the Army for Operations Research Room 2E261, Pentagon Washington, D.C. 20310	2
4.	Commander, United States Army Military Personnel Center Attn: DAPC-PLF 200 Stovall Street Alexandria, Virginia 22332-0400	15
5.	LTC Jack B. Gafford, Code 55GF Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	2
6.	PROF Donald, R. Barr, Code 55BN Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	2
7.	CPT Ronald P. Higham 11 Prospect Avenue Middletown, New York 10940	2

Thesis
H52821 Higham
c.1 A multiple linear regression model for predicting zone A retention by military occupational specialty.